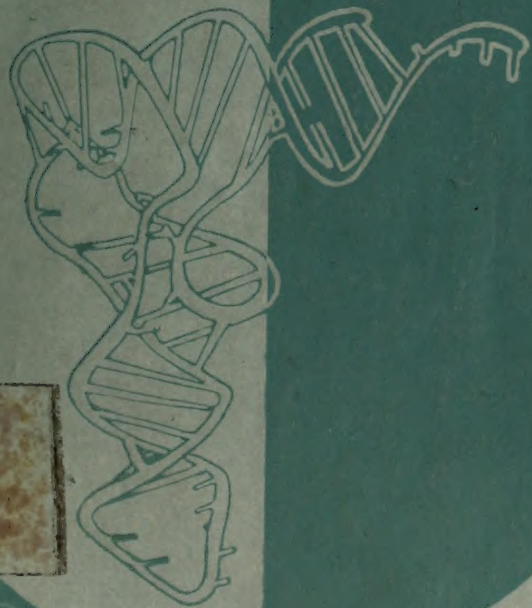
生物化学丛书

核　酸

结构、功能与合成（上册）

王德宝　祁国荣 主编



科学出版社

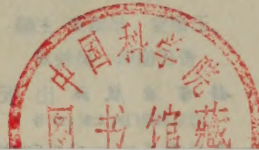
生物化学丛书

核 酸

结构、功能与合成

上 册

王德宝 祁国荣 主编





科学出版社
1986

内 容 简 介

本书根据中国科学院生物化学研究所于1979年和1983年举办的大型生化训练班教材补充修改而成。内容包括核酸的结构（一级结构和高级结构）、主要生物功能（复制、转录和翻译）与人工合成（化学合成和酶促合成）的基础知识及新进展，同时也介绍近年来研究较多的具有生物活性的核苷酸类化合物以及一些较新的核酸分离纯化方法。分上下两册出版。本册主要内容是核酸的成分和结构，DNA的复制、转录与加工等。本书可供生物化学、分子生物学、分子遗传学研究工作者和高等院校包括研究生在内的有关专业师生参考。

生物化学丛书
核 酸
结构、功能与合成
上 册
王德宝 祁国荣 主编
责任编辑 吴铁双
科学出版社出版
北京朝阳门内大街137号
中国科学院印刷厂印刷
新华书店北京发行所发行 各地新华书店经售
*
1986年11月第一版 开本：787×1092 1/16
1986年11月第一次印刷 印张：12 5/8
印数：0001—3,150 字数：276,000
统一书号：13031·3334
本社书号：4055·13—10
定价：3.00元

序　　言

生物化学研究所举办的生化训练班始自 1950 年，开始时着重实验室训练，我们的目的是为了使青年生化工作者掌握这门学科研究的一些新方法、新技术。以后随着我所高研人员的增多，又添加了讲课内容，围绕当时的生化生长点，系统地讲述国际生化的新进展。课文内容逐渐发展充实，并结合在上海科技大学讲授高级生化课的需要，编写成高级生化训练班讲义。1960 年中国科学院在上海召开第一次全国生化学术会议，会议期间各地代表纷纷要求生化所举办一次大型高级生化训练班，为全国各有关单位培训生化人才。因此，生化所于 1961 年举办了一次有四百多学员参加的高级生化训练班，以系统介绍生化学科新知识为主，部分学员并参加了实验训练。十几年来，我们发现通过该次训练班学习的学员，大部分已成为各有关单位的生化科研或教学骨干。这个发现给予生化所同志以极大鼓舞。"文化大革命"中，高级生化训练班横遭批判，但 1972 年以后，各方面仍不断有呼声，要求生化所再次举办高级生化训练班。1976 年我们在所内作一次小型尝试，着重发挥部分中级科研人员在教学中的作用，包括从编写讲义到讲课。1979 年在中国科学院一局的催促和支持下，为克服住宿的困难，我们再次在沪杭两地同时举行一次大型高级生化训练班，人数近五百人，课程内容大为扩充，包括十余年来进展最迅速的生化或分子生物学领域，如分子遗传、DNA 重组、生物膜、免疫生化等等。为适应国内广大生化工作者的需要，特将上述讲课内容整理成书，分册付印，定名为"生物化学丛书"。

本丛书已经出版了三分册，即《蛋白质化学》、《酶的作用原理》和《生物膜的结构与功能》。在此分册付印之前，生化所又于 1983 年在上海科技大学和第二军医大学举办了大型高级生化训练班，因此本分册吸取一些近几年内核酸领域中的重大进展以充实内容，如我国酵母丙氨酸转移核糖核酸的人工合成，DNA 顺序测定方法的新进展以及 DNA 结构的一些新发现等等。今后，随着本门学科领域的发展，本丛书亦将尽量补充新内容，并望国内同行对本书内容不吝批评指正，供今后再版时修改参考。

王 应 睐

目　　录

前言　核酸的历史和文献简介

王　德　宝

十八世纪后期从鸟粪和人尿中发现尿酸，这是世界上发现的第一个嘌呤化合物，十九世纪又相继从膀胱结石中找到黄嘌呤，从鸟粪中找到鸟(便)嘌呤，从牛脾找到次黄嘌呤，这些都是发现核酸以前的一些重要发现，当然，嘌呤究竟具有怎样的结构，上述几种嘌呤化合物有什么生物意义或功能，在当时是不清楚的。

差不多与此同时，生物学家们发现了细胞核，并且认识到细胞是生物体的基本组成单位，是由细胞核及其周围的细胞质组成。

1868—1869 年瑞士 Basel 青年科学家 Friedrich Miescher (1844—1895) 到德国 Tübingen，跟 F. Hoppe-Seyler 学习，研究细胞核的化学成分。他以实验室附近的外科医院收集到的绷带为材料，把脓细胞从绷带上洗下来，先用胃酶盐酸混合液消化这些脓细胞，再用乙醚抽提，细胞核留在水相，沉在器皿底部，可用过滤的方法收集。从这样得到的细胞核中，他提取得到一类化合物，呈酸性，不溶于稀酸，但能溶于稀碱，而且含有相当量的磷元素。这一点引起了 Miescher 的重视，因为在当时从动物组织中得到的含磷化合物只有卵磷脂一种，也是 Hoppe-Seyler 实验室的人发现的。Miescher 把他新发现的化合物命名为"核质"(Nuclein)，并将他的实验结果写成论文送给 Hoppe-Seyler。为了慎重起见，Hoppe-Seyler 没有马上发表，而坚持要亲自重复 Miescher 的实验，等得到同样的结果后，才把 Miescher 的论文和他自己重复的结果在他所主编的杂志上同时发表了。Miescher 发现的"核质"，实际是含有蛋白质的核酸制品，以后 R. Altmann 制得不含蛋白质的核酸，并首先使用"核酸"命名他的制品，这是 Miescher 发现核酸二十年以后的事了。

自此以后，核酸吸引了不少著名化学家的注意，他们对核酸的组成成分感到兴趣，用各种方法分解核酸，并从中分离得到嘌呤碱、嘧啶碱、核糖、脱氧核糖等。其中贡献最大的是 A. Kossel 和他的学生们，以及 P. A. Levene 和他的同事。Kossel 等发现了除鸟嘌呤外(核酸中含有的鸟嘌呤是 J. Piccard 发现的)核酸中常见的其他四种碱基，即腺嘌呤、尿嘧啶、胞嘧啶和胸腺嘧啶。Levene 等则对核糖、2-脱氧核糖、核苷和核苷酸的鉴定等作出了贡献。

但是核酸究竟具有怎样的结构，大小如何，在生物体内起什么作用，有什么功能，直到本世纪初仍然不知道。甚至从事核酸化学组成研究了几十年的 Kossel 也认为核酸没有多大的重要性，因而于1905年改行研究细胞核中的碱性蛋白质。不久第一次世界大战发生，当时世界核酸研究中心的德国是参战的主要一方，核酸的研究当然受到影响而停顿下来了。

直到本世纪二十年代，核酸的主要来源是胸腺和酵母。从胸腺得到的核酸含胸腺嘧啶，从酵母制得的核酸则不含胸腺嘧啶而含尿嘧啶，其他三种碱基，即腺嘌呤、鸟嘌呤和胞嘧啶则两者都有。另外还有一个差别，就是从胸腺核酸得到2-脱氧核糖，而酵母核酸则含有核糖。限于当时的技术水平，从其他动物组织制得的核酸都类似胸腺核酸，而从麦胚这样的植物来源的则和酵母核酸相同，因此一度以为核酸可以分为动物核酸和植物核酸两

大类。以后发现无论在动物或植物细胞核内部都可找到胸腺核酸类型的核酸，而在细胞质内却含有象酵母那样的核酸，于是不再有动植物核酸之分，而出现了核糖核酸（RNA 含核糖）和脱氧核糖核酸（DNA 含脱氧核糖）这样的分类，一直沿用至今。最近有人从艾氏腹水瘤中找到含葡萄糖的核酸，叫做葡萄糖核酸（GNA），其功能还不清楚。

无论 RNA 或 DNA 都是由许许多多核苷酸组成的。核苷酸则由含氮有机碱（嘌呤和嘧啶）、戊糖（核糖或脱氧核糖）和磷酸三种成分构成。核苷酸和核苷酸之间是由磷酸二酯（磷酸和相邻两核苷酸的戊糖形成的）键连接起来的。由于当时实验条件的限制，核酸分子的大小是不清楚的，但是从分析结果看，RNA 或 DNA 都含有四种碱基，它们的克分子[1]比例接近于 1:1:1:1，因此出现了"四核苷酸"的说法，认为核酸就是四核苷酸或者四核苷酸的重复。现在看来四核苷酸的说法当然是错误的，简单而明显的理由是：最小的核酸的分子量也有几万道尔顿，不可能是四核苷酸；核酸中四种常见碱基的克分子比不是 1:1:1:1。

第一次世界大战结束，德国战败了，核酸研究随着分散到世界其他国家，特别是美国和西欧。二十世纪 30 年代又发生第二次世界大战，核酸研究工作又受到了严重冲击而进展迟缓。然而就在这次大战结束前一个很重要的工作发表了，它对核酸的重要生物功能做了最直接的阐明，推动了近四十年来深入广泛的研究，使核酸研究成为生物化学和分子生物学领域内最引人注目的一个分支。1944 年 O. T. Avery、C. M. Macleod 和 M. McCarty 在研究肺炎球菌的转化现象时，发现从一种类型的肺炎球菌制备得到的 DNA 可以引到另一类型的肺炎球菌体内，并使后者的遗传性状改变成前者。他们得出结论遗传物质是 DNA。但由于当时流行的观点：蛋白质（而不是核酸）是遗传的物质基础，Avery 等的实验结果曾一度被怀疑，认为他们的 DNA 制剂内杂有蛋白质。然而 Avevy 等的工作以令人信服的证据证明，DNA 而不是蛋白质才是遗传物质，改变了当时的流行看法，从此核酸的重要性受到了广泛的注意。

1953 年核酸研究的又一个重要里程碑出现了，这就是 J. D. Watson 和 F. H. C. Crick 的 DNA 双螺旋结构学说。这个学说不但阐明了 DNA 的结构，并且为一个 DNA 分子如何复制成两个相同结构的 DNA 分子以及 DNA 怎样传递生物体的遗传性状提供了合理的说明。这项工作为现代分子生物学的奠定起到了关键性的作用，受到了科学界的普遍重视。

自此以后的三十年，核酸研究进展之快，涉及范围之广，影响之大，内容之丰富，在生物科学领域内是少见的。大体说来有下面这些工作。

(1) 核酸的生物合成，包括 DNA 的复制，DNA 转录成 RNA 和转录后的加工以及 RNA 反转录成 DNA 等。

(2) DNA 不仅存在于细胞核内，在核外的一些细胞器中也找到，如真核生物的线粒体、叶绿体等，原核生物的质粒等。RNA 病毒和噬菌体，DNA 病毒和噬菌体的发现。

(3) RNA 和蛋白质生物合成的关系，tRNA，mRNA 和 rRNA 的发现。三联密码的确定，蛋白质生物合成的过程等。

(4) 核酸的新陈代谢，嘌呤和嘧啶核苷酸的生物合成。

1) 1 克分子 (M) = 1 mol；1 克分子浓度 (*M*) = 1 mol/l。

(5) 重要核酸工具酶的相继发现，如核酸酶、聚合酶、限制性内切酶、连接酶等。

(6) 核酸分子中核苷酸排列顺序(一级结构)的测定。

(7) 核酸高级结构的测定。

(8) 基因表达和调节控制的机理。

(9) 真核生物 DNA 和 RNA 与原核生物 DNA 和 RNA 的差异，真核生物结构基因的不连续性，剪接现象的发现等。

(10) 修饰核苷酸和具有重要生物功能小分子活性核苷酸的发现。

(11) 不同核酸之间以及核酸和蛋白质之间的相互作用。核酸蛋白质复合体(如染色体、核糖体、复制体、病毒、噬菌体等)结构的研究。

(12) 细胞核内和其他小分子 RNA 的结构和特殊功能。

(13) 核酸的人工合成。

(14) 综合利用上述核酸研究的成就，衍生出了分子遗传学和基因工程，为医学、农业、工业、环境保护等打开了新局面。

为了说明三十年来核酸研究的大体进展过程和科学界对核酸研究成果的重视程度，现在把近三十年来因从事这方面的研究而获得诺贝尔奖金的科学家列表如下:

科学家	年份	种类	成果
A. R. Tood (英)	1957	化学	确定核苷酸结构，合成二核苷酸等
G. W. Beadle (美) E. L. Tatum (美)	1958 1958	生理学、医学 生理学、医学	化学试剂对基因的控制和影响
J. Lederberg (美)	1958	生理学、医学	提出新的遗传论点
S. Ochoa (美)	1959	生理学、医学	酶促合成核糖多核苷酸
A. Kornberg (美)	1959	生理学、医学	酶促合成 DNA
J. D. Watson (美) F. H. C. Crick (英)	1962 1962	生理学、医学 生理学、医学	DNA 的双螺旋结构
M. H. F. Wilkins (英)	1962	生理学、医学	DNA 的X射线衍射研究
F. Jacob (法) A. M. Lwoff (法) J. L. Monod (法)	1965 1965 1965	生理学、医学 生理学、医学 生理学、医学	基因对酶和病毒合成的控制
R. W. Holley (美)	1968	生理学、医学	酵母 tRNA^{Ala} 一级结构测定
H. G. Khorana (美)	1968	生理学、医学	合成遗传密码
M. W. Nirenberg (美)	1968	生理学、医学	发现遗传密码
M. Delbrück (美) A. D. Hershey (美) S. E. Luria (美)	1969 1969 1969	生理学、医学 生理学、医学 生理学、医学	基因结构和病毒复制机制
E. W. Sutherland (美)	1971	生理学、医学	发现 3',5'-环AMP 和激素作用机制
H. M. Temin (美) D. Baltimore (美)	1975 1975	生理学、医学 生理学、医学	发现反转录酶
R. Dulbecco (意)	1975	生理学、医学	肿瘤病毒和细胞遗传物质之间的相互作用

续表

科学家	年份	种类	成果
W. Arber (瑞士)	1978	生理学、医学	发现细菌限制性内切酶
H. O. Smith (美)	1978	生理学、医学	发现限制性内切酶作用方式的特点
D. Nathans (美)	1978	生理学、医学	用限制性内切酶制成肿瘤病毒的基因图谱
P. Berg (美)	1981	化学	建立 DNA 重组技术
W. Gilbert (美)	1981	化学	DNA 一级结构测定方法
F. Sanger (英)	1981	化学	DNA 一级结构测定方法
A. Klug (英)	1982	化学	建立晶体电子显微技术测定核酸-蛋白质复合体的构造

此外在早期获得诺贝尔奖金者中，和核酸及遗传有关的还有 A. Kossel (德国人，1910 年生理学、医学奖，核酸成分的性质结构等)，T. H. Morgan (美国人，1933 年生理学、医学奖，发表基因论)和 H. J. Muller (美国人，1946 年生理学、医学奖，发现X射线能促使基因变异)等，这种情况在其他科学领域内是罕见的。

几十年来深入广泛的研究，大量资料的积累，核酸的文献可以说是汗牛充栋。下面简单介绍一下这方面的情况，供有志从事核酸研究的同志们参考，自然只能选择一些有代表性或个人认为比较重要的书籍和期刊(列到 1983 年止)，不可能全部罗列出来。

I　书籍

(1) P. A. Levene and L. W. Bass, Nucleic Acids, Chemical Catalog Company, New York, 1931. 这本书的作者 Levene 是早期从事核酸研究最有成就的人之一，对发现核酸以来六十年的研究历程有较详尽而亲切的叙述。

(2) E. Chargaff and J. N. Davidson (eds.) The Nucleic Acids, Academic Press, Inc, New York, Vol.1, 1955; Vol. 2 1955; Vol.3 1960. 这是本世纪六十年代以前收集到的有关核酸化学和生物学材料最丰富的一部书，全书共四十章，每章均由专家撰写。

(3) J. N. Davidson, The Biochemistry of the Nucleic Acids, Chapman & Hall and Science Paperbacks. 此书已出九版，第一版 1950 年发行，至第七版(1972)均由 Davidson编著。Davidson 去世后第八版 (1976 年)起是由他的同事和学生修订的，第九版于 1981年出版。

(4) A. Kornberg, DNA Replication, W. H. Freeman and Company, San Francisco, 2nd ed., 1980; Supplement, 1982.

(5) J. D. Watson, Molecular Biology of the Gene, Benjamin/Cummings Menlo Park, California, 3rd ed., 1976.

(6) J. D. Watson and J. Tooze, The DNA Story, W. H. Freeman and Company, San Francisco, 1981.

(7) P. R. Stewart and D. S. Letham (eds.) The Ribonucleic Acids, Springer-Verlag, New York, 2nd ed. 1977.

(8) P. R. Schimmel, D. Söll and J. N. Abelson (eds) Transfer RNA. Structure, Properties and Recognition, Cold Spring Harbor Laboratory, New York, 1979.

(9) D. Söll, J. N. Abelson and P. R. Schimmel (eds.) Transfer RNA. Biological Aspects. Cold Spring Harbor Laboratory, New York, 1980.

(10) Kivie Moldave (ed.) RNA and Protein Synthesis, Academic Press, Inc, New York, 1981. 此书是从"Methods in Enzymology"丛书中收集有关文章编成的。

(11) G. D. Fasman, Handbook of Biochemistry and Molecular Biology, Nucleic Acids, Vols. I & II, 3rd. ed., 1975, CRC Press, Cleveland.

II 丛书

(1) J. N. Davidson and W. E. Cohn (eds.) Progress in Nucleic Acid Research and Molecular Biology, Academic Press, Inc., New York. 1963 年起出第一卷，第一、二卷原名"Progress in Nucleic Acid Research", 从第三卷起改现名，Davidson 去世后，从第十四卷起由 Cohn 单独主编(第十九卷为 mRNA 专集，由 Cohn 和 E. Volkin 合编)。

(2) S. P. Colowick and N. O. Kaplan (eds.) Methods in Enzymology, Academic Press, Inc., New York, 1955 年起出第一卷，至 1982 年底止已出九十卷。其中第十二卷分 Part A 和 Part B 二册，第二十卷 (Part C)，第二十一卷 (Part D)，第二十九卷(Part E)，第三十卷 (Part F)，第五十九卷 (Part G) 第六十卷 (Part H)，第六十五卷 (Part I) 为核酸或核酸和蛋白质合成专册。第五十一卷为嘌呤和嘧啶核苷酸新陈代谢专册。第六十八卷为重组 DNA 专册。

(3) Cold Spring Harbor Laboratory, Symposium on Quantitative Biology, 1933 年出第一卷，除第二次世界大战期间停出三年外，每年一卷，都是前一年讨论会中报告的论文，其中许多卷是有关核酸、分子遗传、病毒等的专著。

(4) ICN-UCLA Symposium on Molecular and Cellular Biology, Academic Press, Inc., New York. 1972 年起出第一卷。

(5) J. M. Luck 等 (eds.) Annual Review of Biochemistry, Annual Reviews, Inc., Palo Alto California, 1932 年起出第一卷。

(6) C. E. Clifton 等 (eds.) Annual Review of Microbiology, 1974 年起出第一卷。

(7) H. L. Roman 等 (eds.) Annual Review of Genetics, 1967 年起出第一卷。

(8) M. F. Morales 等 (eds.) Annual Review of Biophysics and Bioengineering, 1972 年起出第一卷。第 6—8 卷的出版商同第 5 卷。

III 期刊

没有一种生化期刊不登载核酸研究的文章的，因此世界各国的生化期刊都应该阅览，但由于时间及其他条件的限制，很难或甚至不可能这样做，下面举一些比较重要的期刊供选读。

(1) *Analytical Biochemistry*

(2) *Archives of Biochemistry and Biophysics*

(3) *Biochemical and Biophysical Research Communications*

(4) *Biochemical Genetics*

(5) *Biochemical Journal*

(6) *Biochemistry*

(7) *Biochimica et Biophysica Acta* (1960 年起出核酸专号，1965 年起改为核酸和蛋白

质合成专号，1982 年起又改为基因结构和表达专号)。

(8) *Biochimie*

(9) Биохимия

(10) *Biopolymers*

(11) *Bioscience Reports*

(12) *Cell*

(13) *European Journal of Biochemistry*

(14) *Federation of European Biochemical Societies Letters*

(15) *Federation Proceedings*

(16) *Gene*

(17) *Hoppe-Seyler's Zeitschrift für Physiologische Chemie*

(18) *Journal of American Chemical Society*

(19) *Journal of Bacteriology*

(20) *Journal of Biochemistry*

(21) *Journal of Biological Chemistry*

(22) *Journal of Molecular Biology*

(23) *Journal of Virology*

(24) *Microbiological Reviews*

(25) *Molecular Biology Reports*

(26) *Molecular and General Geneties*

(27) *Nature*

(28) *Nucleic Acids Research*

(29) *Plasmid*

(30) *Proceedings of the National Academy of Sciences, U. S. A.*

(31) *Progress in Biophysics and Molecular Biology*

(32) *Science*

(33) *Scientific American*

(34) *Trends in Biochemical Sciences*

(35) *Virology* 等。

我国出版的有关期刊有

(1) 中国科学 B 辑

(2) 生物化学与生物物理学报

(3) 生物化学与生物物理进展

(4) 生命的化学

(5) 生物科学动态

(6) 实验生物学报

(7) 细胞生物学杂志

(8) 科学通报

(9) 遗传学报

（10）微生物学报等

中国科学和科学通报同时出外文版。

IV　文摘

有关核酸的文章实在太多，单看上面几十种期刊还是不免遗漏（有时甚至是重要的遗漏），因此还要查看文摘，以弥补这个缺陷。

（1）Biochemical Abstracts, Part Nucleic Acids 1971 年起出版，原名“Nucleic Acids Abstracts”，1980 年起改现名。

（2）Biological Abstracts, 1926 年底起出版。

（3）Chemical Abstracts, 1907 年起出版，是目前世界上化学方面最大最全的一部文摘。

（4）Genetics Abstracts, 1968 年起出版。

（5）Virology Abstracts, 1967 年起出版。

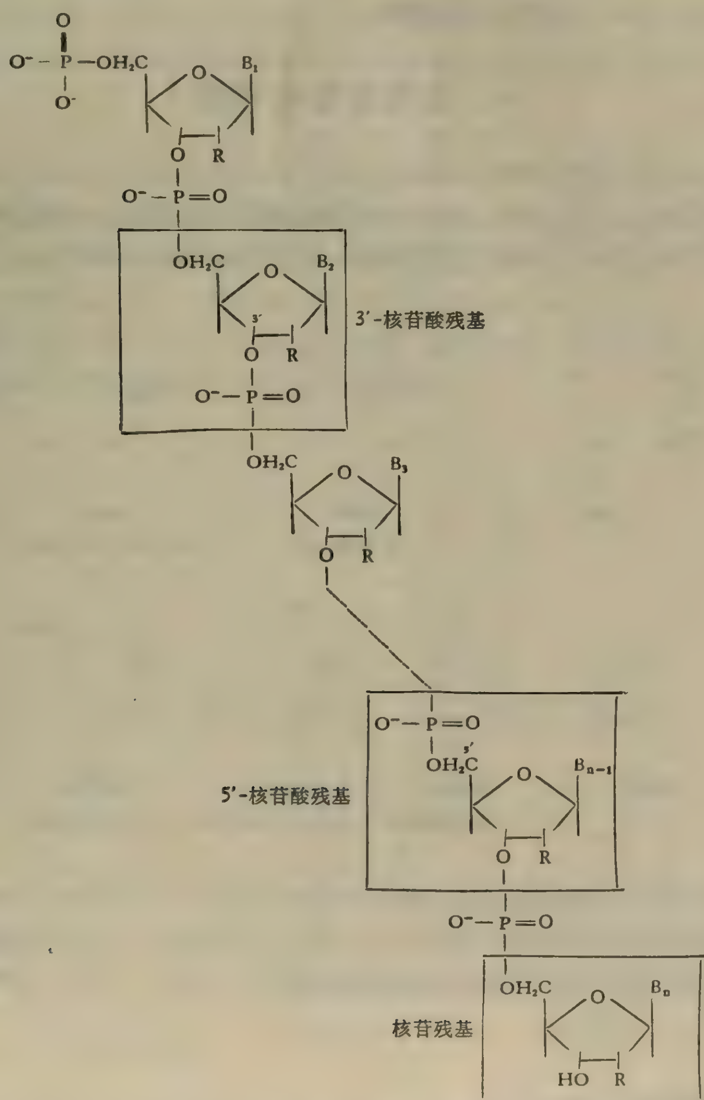
我国的核酸研究起步较晚，到本世纪五十年代中期才开始，由于种种原因，研究的广度和深度除少数工作外，和世界先进水平有不小的差距。这种情况当然要反映到核酸的书籍方面来，建国以来虽然也出版过几种外文译本，但到现在还没有我国自己编写的比较系统的核酸书，不无遗憾。为了弥补这个空白，我们组织我所部分同志，编写了这样一本书。由于本书是生物化学丛书的一部分，在这套丛书里还有分子遗传学和基因工程、新陈代谢等分册，因此有些内容如基因的表达和调控，染色体外的 DNA，病毒和噬菌体的 DNA，DNA 重组，基因工程等将在分子遗传学和基因工程一书中介绍，而核酸的新陈代谢，核苷酸的生物合成等则在新陈代谢书中叙述。

这本书经过许多同志的共同努力写成了，本来应该高兴，但由于下面一些原因又感到不足。1. 作者不是一个人，因此带来一些问题，如文字风格和表达方式不一致，内容难免重复，又可能有遗漏。2. 作者们都不是教师，又缺少教书经验，不善于对复杂的问题作深入浅出的阐述。3. 这本书是为高级生化训练班写的，虽然过去也办过这样的训练班，曾写过一些讲义，但核酸的研究进展很快，可说是日新月异，几年一过，就须重写。这次的改写比较匆忙，而且写书的同志手头都有其他工作，不可能花很多时间脱产写书。4. 更主要的是我们的学术水平有限。因此这本书的选材是否妥当，繁简是否合适，有无重大错误等等，我诚恳地请求广大读者，包括大专学校的老师和同学、科研单位和生产单位的同志们，多提宝贵意见，指出缺点和错误，以便今后重版时加以改正，不胜企盼之至。

第一章　核酸的成分

祁　国　荣

核酸的组成单位是核苷酸，核苷酸由含氮有机碱（碱基）、戊糖和磷酸三者构成。核苷酸组成核酸链的模式如图示：



图中 B_1、B_2、B_3··· B_{n-1}、B_n 代表任一相同或不同的碱基；若R是 OH 则戊糖为核糖，

由它组成的核酸为 RNA。若R是H则戊糖为脱氧核糖，构成 DNA，此外在某些 RNA 中，少数R可以是 OCH_3; 图中方框分别代表3′-核苷酸(由碱基、戊糖和3′位上磷酸根构成)、5′-核苷酸(磷酸根连接在戊糖的5′位上)和核苷(只由碱基和戊糖组成)残基。

核酸的成分早在上世纪末和本世纪初已经确定(见前言)。即 RNA 由腺苷酸、鸟苷酸、胞苷酸和尿苷酸组成，而 DNA 组成分为脱氧腺苷酸、脱氧鸟苷酸、脱氧胞苷酸和脱氧胸苷酸。但是近一、二十年来在不同核酸分子中陆续发现为数众多的修饰成分 (modified component)[1)]，据不完全统计有近八十种。所谓修饰成分就是除了上述八种主要成分外，存在于核酸中含量较少或很少的成分，它们绝大部分是主要成分的衍生物，有的修饰成分的结构非常复杂，则被叫做高度修饰成分 (hypermodified component)。

核酸中的修饰成分与主要成分的连接方式完全相同， 即它们的糖苷键均为β-构型，也都是通过3′, 5′-磷酸二酯键联成核酸链。

一、含氮有机碱(碱基)

组成核酸的碱基主要为嘌呤和嘧啶衍生物，无论在 DNA 或 RNA 中主要嘌呤衍生物为腺嘌呤和鸟嘌呤。在 DNA 中两个主要嘧啶衍生物为胞嘧啶和胸腺嘧啶，尿嘧啶代替胸腺嘧啶则构成 RNA 中两个主要嘧啶碱，它们的结构式分别如下:



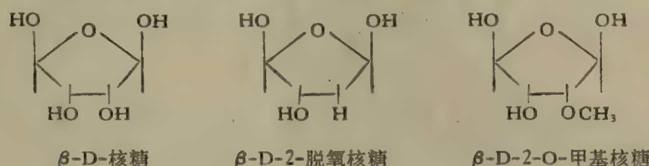
此外，还发现有几十种修饰碱基(常称稀有碱基)，绝大部分为嘌呤、嘧啶化合物，但有二类修饰碱基的母核不是嘌呤、嘧啶环，一类是7-去氮嘌呤，另一类是由鸟嘌呤衍生出来的三杂环母核。所有已发现的修饰碱基的结构将在“核苷”一节中一并讨论(结构式见附录一)。

---

1) 修饰成分以前常被叫做稀有成分或附加成分，英文曾用名有: minor. unusual, additional, rare, odd component 等。

二、戊　糖

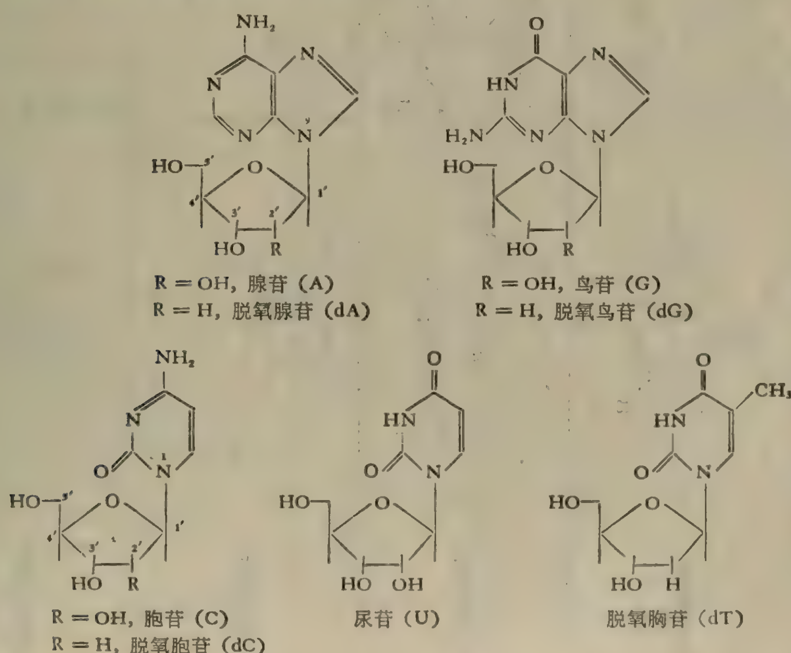
构成 RNA 链的戊糖是 D-核糖，DNA 链中的戊糖是 D-2-脱氧核糖。在某些 RNA 中发现有部分 D-2-O-甲基核糖。所有这三种糖与碱基连接都是 β-构型。



在核酸中还发现有几个己糖，如 RNA 中的甘露糖、半乳糖，DNA 中的葡萄糖．甘露糖等，但这些己糖并不组成核酸的骨架，而是与碱基的侧链相联接(见"核苷"一节)。

三、核　苷

由碱基与核糖化合物形成的 β-糖苷统称核苷。核酸的戊糖组分有三种，故核苷可分为核糖核苷(简称核苷)、脱氧核糖核苷(简称脱氧核苷)和 2′-O-甲基核苷。在核苷分子中，糖环上的原子编号是 1′、2′、3′、4′、5′，以区别于嘌呤、嘧啶环上的 1、2、3、4、5、……。核糖核苷主要有四种：腺苷、鸟苷、胞苷和尿苷。脱氧核糖核苷主要也是四种：脱氧腺苷、脱氧鸟苷、脱氧胞苷和脱氧胸苷，结构式分别如下：



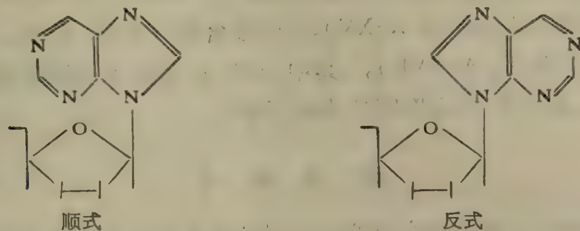
按照碱基的区别又可分为嘌呤核苷和嘧啶核苷，前者通过 $N_9—C_{1'}$ 连接，后者则是$N_1—C_{1'}$ 相连。N-糖苷键对酸是不稳定的。虽然理论上嘌呤或嘧啶环形成的 N-糖苷键可自

表 1-1 不同 pH 条件下核苷的解离态*

核　苷	pH 2	pH 7	pH 12
腺苷或脱氧腺苷			
鸟苷或脱氧鸟苷			
胞苷或脱氧胞苷			
尿苷或脱氧胸苷			

* R 为核糖或脱氧核糖；R′ 为H或 CH_3。

由转动，但事实上由于空间障碍限制其自由转动；在天然核苷中，反式（anti-form）构象较顺式（syn-form）更为合适，在 DNA 双螺旋中碱基配对也是以反式定位的。顺式构象的写出方式仅仅是习惯上的方便。



顺式　　　　反式

在核苷中，杂环上的氨基或酮基可以互变为亚氨基或烯醇基，它们在不同 pH 条件下的解离态见表 1-1。

核酸中的修饰成分大部分是以修饰核苷产物被分离鉴定的。修饰核苷可分为三类：(1)由修饰碱基组成的核苷，(2)由 2-O-甲基核糖组成的核苷，(3)碱基与戊糖连接方式与众不同的核苷。其中以第 1 类发现最多，第 3 类肯定结构的只有假尿苷（ϕ）及其衍生物。据1982 年不完全统计，RNA 修饰核苷共 69 个，而 DNA 中只有 9 个。RNA 中修饰成分绝大部分是在 tRNA 中发现的，而 DNA 中修饰成分主要由噬菌体 DNA 分离得到。

RNA 中修饰核苷可区分为修饰腺苷、修饰鸟苷、修饰胞苷、修饰尿苷和 2′-O-甲基核苷，以下分别进行讨论。

表 1-2　RNA 中的修饰腺苷

腺苷衍生物*	代　号	符　号	RNA 中分布**				发现年份
			t	r	m	sn	
1-甲基腺苷	A-I	m^1A	+				1960
2-甲基腺苷	A-II	m^2A	+	+			1958
N^6-甲基腺苷	A-III	m^6A	+	+	+	+	1956
N^6，N^6-二甲基腺苷	A-IV	m^6_2A	+	+			1958
2-甲硫基腺苷	A-V	ms^2A	+				1973
N^6-甘氨酸羰基腺苷(+)	A-VI	gc^6A	+				1970
N^6-苏氨酸羰基腺苷(+)	A-VII	tc^6A	+				1969
N^6-甲基-N^6-苏氨酸羰基腺苷(+)	A-VIII	m^6tc^6A	+				1972
2-甲硫基-N^6-苏氨酸羰基腺苷(+)	A-IX	ms^2tc^6A	+				1981
N^6-三羟甲基甲氨基苏氨酸羰基腺苷(+)	A-X	(N^+)	+				1976
N^6-(Δ^2-异戊烯)-腺苷	A-XI	i^6A	+				1966
2-甲硫基-N^6-(Δ^2-异戊烯)-腺苷	A-XII	ms^2i^6A	+				1970
N^6-(顺-4-羟异戊烯)-腺苷	A-XIII	oi^6A（顺式）	+				1972
N^6-(反-4-羟异戊烯)-腺苷	A-XIV	oi^6A（反式）	+				1967
2-甲硫基-N^6-(顺-4-羟异戊烯)-腺苷	A-XV	ms^2oi^6A（顺式）	+				1970
肌苷(次黄嘌呤核苷)	A-XVI	I	+				1963
1-甲基肌苷	A-XVII	m^1I	+				1963

* 结构式见附录一，有“+”号的不是标准命名。

** t，r，m，sn 分别代表 tRNA，rRNA，mRNA（包括病毒 RNA）和 snRNA（细胞核小分子 RNA）。

修 饰 腺 苷

表 1-2 列出 RNA 中发现的十七种腺苷衍生物及其分布，包括甲基化腺苷（A-I 至 A-IV, A-VIII 和 A-XVII），侧链带氨基酸的腺苷（A-VI 至 A-X），侧链含异戊烯基的细胞激动素类似物（A-XI 至 A-XV），有的含硫（A-V, A-IX, A-XII 和 A-XV），也有脱氨产物（A-XVI 和 A-XVII）。其中 m^1A 在碱性下可互变为 m^6A，酸性处理则嘌呤环开环；tc^6A、m^6tc^6A 和 m^1I 则易为碱所破坏。

修 饰 鸟 苷

虽然Q核苷和 W(Y) 核苷化合物的碱基母核不是嘌呤环，但可以把它们看作是鸟苷的衍生物(见结构式)。表 1-3 列出已知的 12 种修饰鸟苷,主要为甲基化鸟苷 (G-I 至 G-VI)。7-位甲基化鸟苷（G-IV 至 G-VI）易为碱打开咪唑环，然后经酸处理使糖苷键断裂。利用这个性质可以特异地在 m^7G 处切断 tRNA 分子,研究其结构与功能的关系。

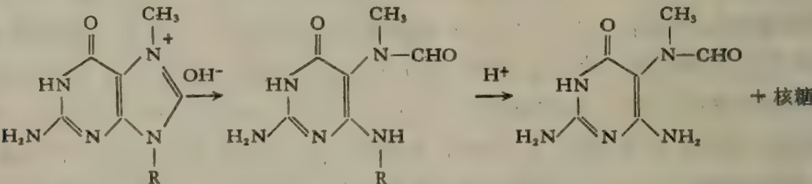


表 1-3 RNA 中的修饰鸟苷*

鸟苷衍生物	代 号	符 号	RNA 中分布				发现年份
			t	r	m	sn	
1-甲基鸟苷	G-I	m^1G	+	+			1958
N^2-甲基鸟苷	G-II	m^2G	+	+		+	1958
N^2, N^2-二甲基鸟苷	G-III	m^2_2G	+				1959
7-甲基鸟苷	G-IV	m^7G	+	+	+		1963
N^2, 7-二甲基鸟苷	G-V	$m^{2,7}_2G$			+		1977
N^2, N^2,7-三甲基鸟苷	G-VI	$m^{2,2,7}_3G$				+	1969
Wyosine	G-VII	W(Yt)	+				1971
Wyobutosine	G-VIII	yW(Y)	+				1968
Peroxywyobutosine	G-IX	oyW (Yr, Yw)	+				1969
Queuosine	G-X	Q	+				1975
Mannosylqueuosine	G-XI	manQ	+				1976
Galactosylqueuosine	G-XII	galQ	+				1976

* 注释见表 1-2。

1968 年从酵母 $tRNA^{Phe}$ 中分离出一种萤光核苷Y,其后在其他来源 $tRNA^{Phe}$ 中也发现Y的类似物,依其来源分别称为 Yt、Yw 和 Yr(小写字母分别标明来自圆酵母、小麦胚和鼠肝 $tRNA^{Phe}$)，所有这些化合物都含相同的二甲基三杂环部分,学名叫 1, N^2-异丙烯-3-甲基鸟苷,但Y和 Yw 带侧链而 Yt 则无。有人建议将Y改为W，它们的结构和对应关系见表 1-4。W(Y)核苷均位于 $tRNA^{Phe}$ 的反密码子的 3′邻位(见第三章)，它们都是酸碱不稳定化合物,因此可用特异条件在该修饰核苷处切开 $tRNA^{Phe}$ 分子。

表 1-4　W(Y) 核苷的命名和结构

结　　构	代　号	符　　号	命　名*	R′ =
(结构式：H_3C—，R′，O，N，CH_3，R)	G-VII	W，Yt	Wyosine	—H
	G-VIII	yW，Y	Wybutosine	$—CH_2CH_2CH(COOCH_3)(NHCOOCH_3)$
	G-IX	oyW，Yw、Yr	Peroxywybutosine	$—CH_2CH(OOH)CH(COOCH_3)(NHCOOCH_3)$

* 其碱基部分分别叫 Wye，Wybutine 和 Peroxywybutine。

Q核苷也是 1968 年发现的，但 1975 年才确定其结构，碱基骨架为 7-去氮鸟嘌呤，C_7 通过氨甲基与环戊烯二醇相连，有的Q化合物其侧链羟基与糖连接形成 β-D-甘露糖苷 (man Q) 和 β-D-半乳糖苷 (gal Q) (见表 1-5)。Q核苷对酸碱稳定，常出现在某些 tRNA 反密码子的摆动位置上(见第三章)。

表 1-5　Q 核苷的结构

结　　构	代　号	符　　号	R′	R″
(结构式：O，R′，HN，H_2N，N，N，R)	G-X	Q*	CH_2NH—(环戊烯，HO，OH)	
	G-XI	manQ	CH_2NH—(环戊烯，R″O，OH)	(甘露糖：HO，OH，OH，HO，O，H)
	G-XII	galQ		(半乳糖：OH，OH，HO，O，HO，H)

* Queuosine 的符号，其碱基叫 Queuine

修 饰 胞 苷

RNA 中胞苷衍生物种类较少，目前只发现六种，其取代侧链遍布于胞嘧啶环上各个部位(见表 1-6)。 m^3C (C-II) 和 ac^4C (C-IV) 为碱不稳定化合物，前者碱处理后易失去紫外吸收特征。

修 饰 尿 苷

在已发现的修饰核苷中，以尿苷衍生物种类最多，共廿五种 (表 1-7)，其中大部分为

表 1-6 RNA 中的修饰胞苷*

胞苷衍生物	代　号	符　号	RNA 中分布				发现年份
			t	r	m	sn	
2-硫代胞苷	C-I	s^2C	+				1968
3-甲基胞苷	C-II	m^3C	+	+			1963
N^4-甲基胞苷	C-III	m^4C		+			1978
N^4-乙酰胞苷	C-IV	ac^4C	+				1966
5-甲基胞苷	C-V	m^5C	+	+			1958

* 注释见表 1-2。

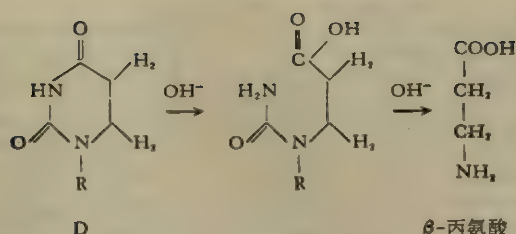
表 1-7 RNA 中的修饰尿苷*

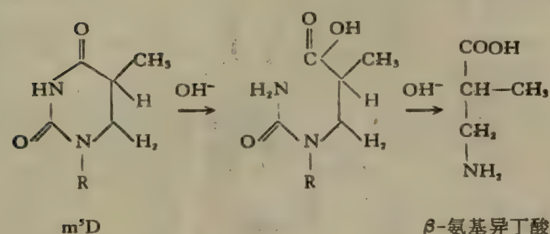
尿苷衍生物	代　号	符　号	RNA 中分布				发现年份
			t	r	m	sn	
2-硫代尿苷	U-I	s^2U	+				1981
3-甲基尿苷	U-II	m^3U	+	+			1963
3-(3-氨基-3-羧丙基)尿苷	U-III	ncp^3U(X)	+				1974
4-硫代尿苷	U-IV	s^4U	+				1965
5-羟基尿苷	U-V	o^5U	+				1966
5-甲氧基尿苷	U-VI	mo^5U	+				1976
5-羧甲氧基尿苷	U-VII	cmo^5U(V)	+				1971
5-甲氧基羰基甲氧基尿苷	U-VIII	$mcmo^5U$(mV)	+				1977
5-甲基尿苷(胸苷)	U-IX	m^5U(T)	+	+			1958
5-甲基-2-硫代尿苷	U-X	m^5s^2U	+				1971
5-羟甲基尿苷	U-XI	om^5U	+				1968
5-羧甲基尿苷	U-XII	cm^5U	+				1967
5-羧甲基-2-硫代尿苷	U-XIII	cm^5s^2U	+				1973
5-(2-羟羧甲基)尿苷	U-XIV	com^5U	+				1979
5-甲氧基羰甲基尿苷	U-XV	mcm^5U	+				1970
5-甲氧基羰甲基-2-硫代尿苷	U-XVI	mcm^5s^2U	+				1968
5-氨甲酰甲基尿苷	U-XVII	ncm^5U	+				1972
5-甲氨甲基-2-硫代尿苷	U-XVIII	mnm^5s^2U	+				1968
5-羧甲基氨甲基尿苷	U-XIX	$cmnm^5U$	+				1978
5-羧甲氨甲基-2-硫代尿苷	U-XX	$cmnm^5s^2U$	+				1981
5,6-二氢尿苷	U-XXI	D(hU)	+				1965
5-甲基-5,6-二氢尿苷**	U-XXII	m^5U					1968
假尿苷	U-XXIII	Ψ	+	+		+	1957
1-甲基假尿苷	U-XXIV	$m^1\Psi$	+				1982
1-甲基-3-(3-氨基-3-羧丙基)-假尿苷	U-XXV	$m^1ncp^3\Psi$ (amΨ)		+			1974

* 见表 1-2 注释。

** 在染色体 RNA 中发现。

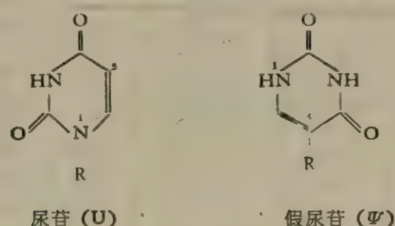
5-位取代化合物 (U-IV 至 U-XX 和 U-XXII)。侧链基团有甲基、羟基、甲氧基、羟甲基、羧甲氧基及其甲酯、羧甲基及其甲酯和酯胺衍生物、氨基酸、羟基酸等，它们具有通常的有机反应性。有的侧链含硫(七种)；有的 5, 6-加氢 (U-XX 和 U-XXII)，加氢后共轭体系被破坏而失去紫外吸收特性，因而早期不易为人们所发现，5, 6-二氢尿苷 (D) 经碱处理得到 β-丙氨酸，5-甲基-5, 6-二氢尿苷 (m^5D) 则转变为 β-氨基异丁酸，故测定其





产物可作为D和 m^5D 的定量分析。m^3U（U-II）与 m^3C（C-II）相似，易为碱所破坏。m^5U（T）（U-IX）为胸腺嘧啶核糖核苷即胸苷。我们知道，RNA 和 DNA 组成中差异之一是前者含尿嘧啶，后者则以胸腺嘧啶代替，随着核酸组成研究的进展，已发现 RNA 中含胸腺嘧啶，有的 DNA 中存在尿嘧啶(见表 1-9)；特别在 tRNA 中，由胸腺嘧啶组成的胸苷酸几乎已是不可缺少的组分。

结构比较奇特的是假尿苷（ψ，U-XXIII）及其衍生物（U-XXIV 和 U-XXV）。这是目前确定的糖苷键连接方式唯一与众不同的核苷，在假尿苷中，其碱基为一般的尿嘧啶，戊糖亦是通常的 D-核糖。但连接方式为 $C_5—C_{1'}$，而不是常规的 $N_1—C_{1'}$，构成 C—C 键而不是 N—C 键，因而产生一系列奇特的物理和化学性质。假尿苷是尿苷的异构体，

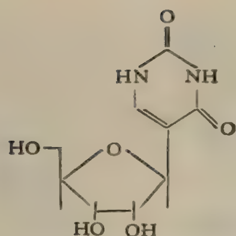


若用 1N HCl 100℃ 处理假尿苷，则通过下列互变机制产生一系列异构体(见下页)。

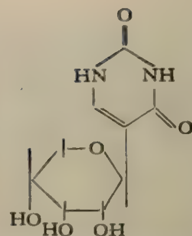
2′-O-甲基核苷

RNA 中含有少量的 D-2-O-甲基核糖代替 D-核糖，由它们构成的核苷称 2′-O-甲基核苷，表 1-8 列出已发现的 2′-O-甲基核苷，其中有几个为碱基和核糖双取代的化合物。

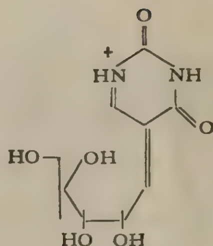
DNA 分子中的修饰核苷，除了 N^6-甲基脱氧腺苷（m^6dA）和 5-甲基脱氧胞苷（m^5dC）较为广泛分布外，其余修饰成分都是在噬菌体 DNA 中发现的(见表 1-9)，其中某些修饰成分的碱基进一步被糖基所取代，故也称作高度修饰成分。如大肠杆菌 T2、T4 和 T6 噬菌体 DNA 含有与 5-羟甲基胞嘧啶中羟基相连的葡萄糖，枯草杆菌 SP8 噬菌

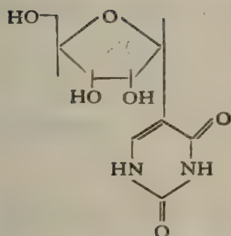


5-β-D-呋喃核糖基尿嘧啶(Ψ)

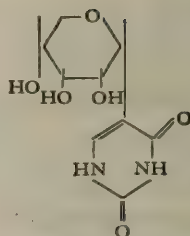


5-β-D-吡喃核糖基尿嘧啶





5-α-D-呋喃核糖基尿嘧啶



5-α-D-吡喃核糖基尿嘧啶

表 1-8 RNA 中的 2′-O-甲基核苷*

2′-O-甲基核苷	符　　号	RNA 中分布				发现年份
		t	r	m	sn	
2′-O-甲基腺苷	Am	+	+	+	+	1959
2′-O-甲基鸟苷	Gm	+	+	+	+	1959
2′-O-甲基胞苷	Cm	+	+	+	+	1959
2′-O-甲基尿苷	Um	+	+	+		1959
2′-O-甲基胸苷	Tm	+				1974
2′-O-甲基假尿苷	ψm	+	+			1964
2′-O-甲基肌苷	Im					1974
N^6, 2′-O-二甲基腺苷	m^6Am			+		1977
N^4, 2′-O-二甲基胞苷	m^4Cm		+			1966
5-羧甲基-2′-O-甲基尿苷	cm^5Um	+				1975
5-氨甲酰甲基-2′-O-甲基尿苷	ncm^5Um	+				1976

* 注释见表 1-2。

体 DNA 也含有葡萄糖，其变种 SP8 Ts 则发现有甘露糖，SP15 噬菌体 DNA 的脱氧尿苷侧链的4-羟基带三个葡萄糖分子。

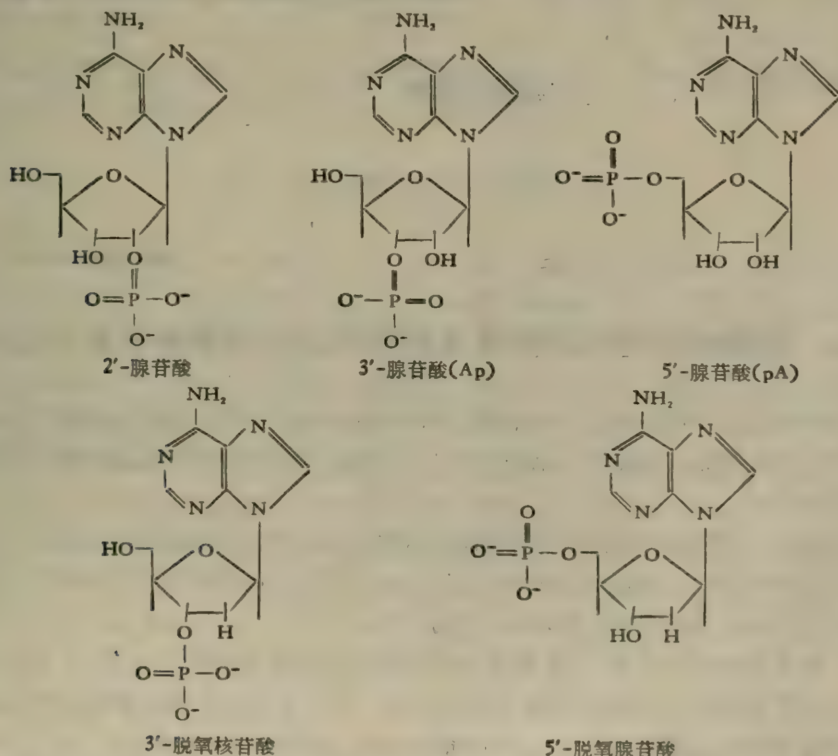
表 1-9 DNA 中的修饰成分*

名称	符号	DNA 来源		发现年份
		噬菌体	宿主	
N^6-甲基脱氧腺苷	m^6dA		*Synechococcus elongatus*	1977
2-氨基脱氧腺苷	n^2dA	S2L		
5-甲基脱氧胞苷	m^5dC	XP12	黄单孢杆菌	1968
5-羟甲基脱氧胞苷	om^5dC	$T_{双数}$	大肠杆菌	1953
脱氧尿苷	dU	PBS2	枯草杆菌	1963
5-羟甲基脱氧尿苷	om^5dU	ϕe	枯草杆菌	1962
5-(4,5-二羟戊基)脱氧尿苷	dhpdU	SP15	枯草杆菌	1969
α-丁二胺脱氧胸苷(+)	putdT	ϕW14	假单孢杆菌	1973
α-谷氨酸脱氧胸苷(+)	gludT	SP10	枯草杆菌	1963 1971

* 结构式见附录一，(+)表示非标准命名。碱基侧链连有糖基的见正文说明。

四、核 苷 酸

核苷的磷酸酯叫核苷酸。核糖有三个游离羟基，所以核糖核苷酸有 2′-、3′-和 5′核苷酸三种，脱氧核糖只有二个羟基可以酯化，故脱氧核糖核苷酸只有 3′-和 5′-核苷酸，2′-O-甲基核苷酸也只有二种。以腺嘌呤核苷酸为例：



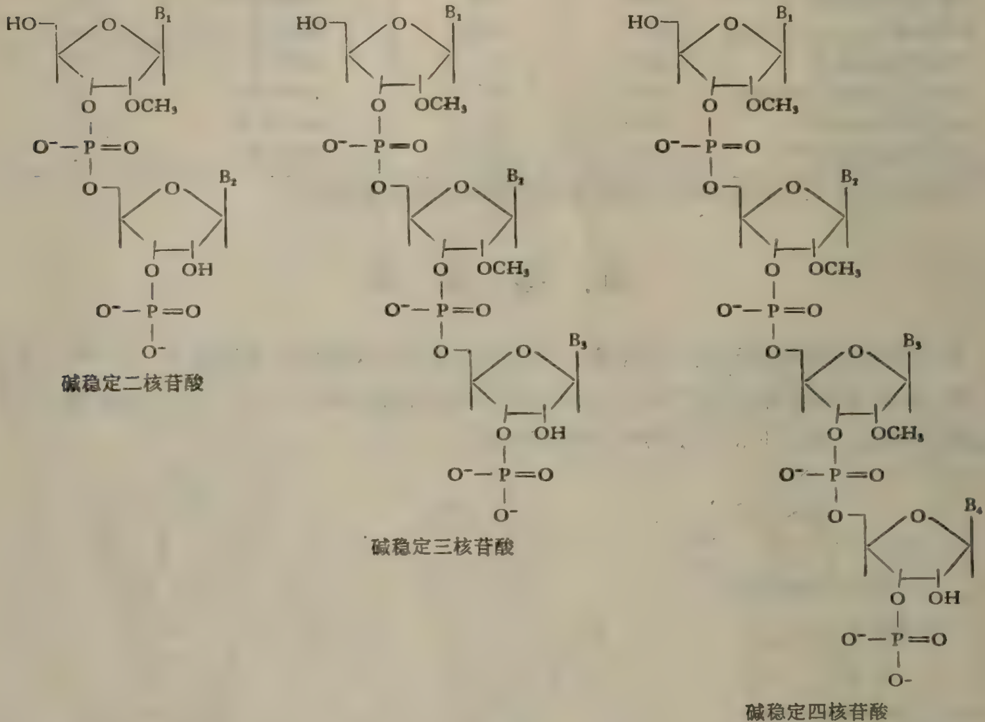
2′-腺苷酸　　3′-腺苷酸(Ap)　　5′-腺苷酸(pA)

3′-脱氧核苷酸　　5′-脱氧腺苷酸

核酸分子是由核苷酸通过 3′→5′ 磷酸二酯键连接而成，因此在核酸链中可以认为只有 3′-和 5′-两种核苷酸，而不存在 2′-核苷酸(见第 1 页图)。但 1978 年发现的一类活性寡核苷酸 pppA (2′) p(5′) A(2′) p(5′)A 中腺苷酸间却是 2′—5′连接(见第九章第六节)，RNA

被碱水解经过环化中间物，则可得到 2′-与 3′-核苷酸混合物。

2′-O-甲基核苷酸构成的 RNA 中磷酸二酯键具有特殊的性质，如不能被碱或 3′-RNase 所水解(因为不能形成 2′, 3′-环化磷酸中间物)，故可以从 RNA 的碱解或酶解产物中分离出各种碱稳定的寡核苷酸。2′-O-甲基核苷酸广泛分布于各种 RNA 中，是真核 mRNA 帽子结构的常见组分。



碱稳定二核苷酸　　碱稳定三核苷酸　　碱稳定四核苷酸

五、核酸成分的缩写符号及核苷酸在核酸链中的表示方式

由于核酸组成分种类繁多，核苷酸在核酸链中的排列顺序及其方向性对于核酸结构及功能表达至关重要，因此统一缩写符号及书写方式是非常必要的。按照目前文献常用表示法归纳如下:

(1) 因为在核酸结构中是以核苷或核苷酸符号表示的，故碱基符号使用不多。为了与核苷符号相区别，碱基一般采用三字符号，即英文名称前三个字母表示，如腺嘌呤(adenine) 为 Ade，胸腺嘧啶 (thymine) 为 Thy。

(2) 核苷用单字符号表示，如腺苷、鸟苷、胞苷和尿苷分别为 A、G、C 和 U; 脱氧核苷则在单字符号前加一小写字母 d，如 dA、dG、dC 和 dT 分别代表脱氧腺苷、脱氧鸟苷、脱氧胞苷和脱氧胸苷。对于由修饰碱基组成的修饰核苷，除部分有独自的单字符号(如 D、I、Q、W 等)外，其余均在 A、G、C、U 和 dA、dG、dC、dT 的左面以小写字母及数码注明其取代基团性质、数目和位置(见表 1-2、1-3、1-6、1-7、1-9 及附录一)，如 m_2^6A 标明腺苷的六位有二个甲基即 N^6, N^6-二甲基腺苷，$m_3^{2,2,7}$G 表示鸟苷的第二位有二个甲

基，第七位有一个甲基，共三个甲基，即 N^2, N^2, 7-三甲基鸟苷；2′-O-甲基核苷的表示是在核苷右方注上小写字母 m，如 Am 代表 2′-O-甲基腺苷，m^4Cm 为 N^4, 2′-O-二甲基胞苷等(见表 1-8)。

(3) 核苷酸的表示方法是在核苷符号的左方小写字母 p 表示 5′-磷酸酯，右方代表 3′-磷酸酯，如 pA 为 5′-腺苷酸。Cp 为 3′-胞苷酸，依次类推。核苷多磷酸的表示法是：ppU 代表 5′-尿苷二磷酸，pppA 为 5′-腺苷三磷酸，即 ATP，ppGpp 表示鸟苷四磷酸，其中二个磷酸根在鸟苷的 5′位，另二个在 3′位。2′-核苷酸无专一表示符号。2′, 3′-环化核苷酸一般用>P 表示，如 U > P 代表 2′,3′-环化尿苷酸，3′, 5′-环化核苷酸一般书写为 cAMP、cGMP 等。

(4) 书写一条核酸链通常从 5′端至 3′端由左向右表示。如(5′)pApGpC···pUpC(3′)，其中的 p 在核苷的右方代表 3′连接，在左下方表示 5′相连。这样两核苷之间正好表示 3′→5′连接。在书写一条互补双键核酸分子时，由于两条链是互为反向的，因此在末端写上 5′和 3′必不可少，如

```
5′G A A T T C 3′
  | | | | | |
3′C T T A A G 5′
```

对于一条长的核酸链，或者从左向右分行书写，或者连续转弯书写均可(如 tRNA 的三叶草结构，见第三章)，这时只要从 5′端开始顺着核酸链往下阅读便是。

为了简便，核酸链中的 p 通常被省略(末端的 p 不可省)，或在核苷间加一小点或一短线，如 (5′) pAGC···UC (3′) 或(5′) pA·G·C······U·C (3′) 或 (5′)pA-G-C···U-C (3′)。若为 2′—5′ 连接，书写时必需注明(见第九章第六节)，5′→5′ 连接也需要标明，如真核生物 mRNA 的帽子结构(见第三章)。以免产生误会。

六、核酸成分的分离鉴定

在 1955 年出版的 E. Chargaff 和 J. N. Davidson 合编的《The Nucleic Acids》(Academic Press, Vol. I)[1)]一书中，曾分章详细地叙述了用离子交换柱层析、纸层析和纸电泳分离核酸组成分的方法。对于核酸中主要成分包括碱基、核苷、核苷酸及多磷酸衍生物的分离，许多后来发展的生化技术如气相层析、反相层析、薄层层析和高压液相层析等都能够达到良好的效果。随着修饰成分的不断发现，1971 年 Hall 在他的专著中较为详细地介绍了当时已发现的修饰核苷的分离方法(文献[1])。由于核酸中修饰成分含量甚微，从百分之几到万分之几，一般需采用灵敏的分离鉴定技术，如同位素标记技术。为了分开尽可能多的核酸成分，常采用快速而微量的双向薄层层析图谱法。附录二给出几个图谱，并且列出其分离条件，包括 20 几种 3′-核苷酸的分离，近 20 种 5′-核苷酸的分离，以及分开 16 种可能排列的碱稳定二核苷酸中的 14 种。1975 年提出一种可以分离近 40 种修饰核苷的双向薄层层析法，是目前分离核酸组成效果最好的一种方法。

鉴定核酸成分常用的方法是在不同 pH 条件下测定其紫外吸收光谱，包括 $\lambda_{最大值}$、$\varepsilon_{最大值}$、$\lambda_{最小值}$和一些特征光吸收比值，这些都是鉴定和定量测定不同核酸组成的重要数据(见文献[2])。一般地说，同一碱基的核苷酸光谱与其核苷差异不大；同样，相同碱基的

1) 有中译本《核酸》(第一卷)，科学出版社，1963，译者：黄德民。

2′-O-甲基核苷或脱氧核苷与其核苷的光谱接近；但是游离碱基光谱与其核苷或脱氧核苷相比则往往差异甚大。例如 pC、Cp、C、dC 和 Cm 光谱性质相似，它们与 Cyt 则有明显的不同。

包括大部分已发现的修饰成分在内的核酸组成分的紫外吸收高峰的波长在 260nm附近。范围常在 250—280nm 之间，低峰波长为 220—240nm 左右。鸟嘌呤、鸟苷及其衍生物的光谱曲线除了常见的吸收高峰外，在红向还有一弯曲部分，称作"肩膀（Shoulder)"。有的修饰成分的紫外吸收光谱较特殊，如 D, s^4U, ac^4C, ms^2i^6A, mnm^5-s^2U 等的紫外吸收高峰远离 260nm（见文献[2]）。

附录三列出以核苷为代表的核酸组成分的紫外吸收性质，以供查阅。

利用萤光性质也是鉴定某些核酸成分的方法之一。鸟嘌呤及其衍生物在某些条件下表示出很高的萤光强度。W核苷也具有很强的萤光性质。一般嘧啶化合物和腺嘌呤衍生物的萤光强度甚低，但某些核酸组分可特异地与萤光试剂结合产生萤光化合物，从而可用萤光法测定这些成分。

要想最后确定一个未知结构的成分，最好与标准样品进行比较，包括光谱性质、层析、电泳行为，以及需要分析的其他性质如 p*K* 值、降解产物鉴定等。如果掌握一些已有的标准分离技术，采用图谱对照方法也可初步判断某些修饰成分。对于结构复杂的高度修饰成分，只有采用新的物化技术鉴定其结构才成为可能，常用方法有质谱、核磁共振光谱、红外光谱等。

除了上述介绍的近 80 种修饰成分外，有关书刊杂志还报道了存在的其他一些修饰成分，但有的证据不足，有的已被否定，有的其结构和性质有待进一步鉴定，故本书未列入。

参考文献

[1] Hall, R. H. 1971. The modified nucleosides in nucleic acids, Columbia University Press.

[2] Dunn, D. B. and Hall, R. H. 1975, Purines, pyrimidines, nucleosides and nucleotides: Physical constant and spectral properties, in "CRC H.ndbook of biochemistry and molecular biology, 3rd ed, Fasman, G. D. (ed.), Chemical Rubber Co. Press, Cleveland." **1**, 65.

[3] Hall, R. H. and Dunn, D. B. 1975. Natural occurrence of the modified nucleoside, ibid, **1**, 216.

[4] Nishimura, S. 1972. Minor component in tRNA: Their characterization, location and function. in "*Prog. Nucleic Acid Res. Mol. Biol., ed.* Davidson, J. N. and Cohn, W. E." **12**, 49.

[5] Feloman, M. Ya., 1977, Minor Components in tRNA: The Location-Function Relationship, in *Prog. Biophys. Mol. Biol.*, **32**, 83—102, eds. Butter, J. A. V. and Noble, D., Pergamon Press.

[6] Agris, P. F.. 1980, The Modified Nucleosides of tRNA: A bibliography of biochemical and biophysical studies from 1970—1979, N. Y. Liss.

[7] Warren, R. A. J. 1980. Modified bases in bacteriophage DNAs in *Ann. Rev. Microbiol. Ann. Rev. Inc. Palo Alto Califonia*, **34** 137.

第二章 DNA 结构

李其樑 徐有成

DNA 是十分巨大的生物高分子。它是由四种脱氧核糖核苷酸通过磷酸二酯键聚合而成的。最小的天然 DNA 分子也包含了几千碱基对，分子量在 10^6 以上。每个人的细胞染色体中含有 10^9 碱基对。表 2-1 列出了若干生物中每个细胞所含 DNA 的碱基对数目。从表中所列的数据可以看出，每个细胞中 DNA 碱基对总数随着生物进化水平的上升而相应增加。但是也存在一些例外的情况，例如南美肺鱼的 DNA 大小约为 10.2×10^{10}碱基对，当然我们并不认为肺鱼的进化水平比人类的水平更高。在植物和两栖类动物中还有一些例外。

表 2-1 DNA 分子和基因组的大小

	碱基(对)数目	总长度 (mm)	形态
病毒:			
多瘤病毒	5.1×10^3	0.0017	环状
SV40	5226	0.0018	环状
M13	6407	0.0019	环状
fd	6408	0.0019	环状
P4	1.5×10^4	0.0051	线状
T7	3.5×10^4	0.012	线状
λ	48502	0.016	线状
T2,T4, T6	1.8×10^5	0.061	线状
细菌:			
枝原体	7.6×10^5	0.26	环状
大肠杆菌	4×10^6	1.36	环状
真核生物(单倍体):			
酵母	1.35×10^7	4.6	分布于 7 对染色体中
果蝇	1.65×10^8	56	分布于 4 对染色体中
人	2.9×10^9	990	分布于 23 对染色体中
南美肺鱼	1.02×10^{11}	34,700	分布于 19 对染色体中

DNA 分子能以线状或环状的形式存在。绝大多数 DNA 分子是由两条碱基互补的单链构成的，只有少数几种 DNA 是单链的，如 φX174，M13 等。

线状双链 DNA 的末端可能是齐头的，也可能带有单链粘性末端，也可能是粘性末端补满以后形成 5′，3′末端重复顺序，也可能具有特殊的"转排"结构。它们之间的相互关系可以通过图 2-1 的例子联系起来。

DNA 分子主要由四种碱基组成，即腺嘌呤、鸟嘌呤、胞嘧啶和胸腺嘧啶，修饰碱基含量很低。真核生物 DNA 中有少量 5-甲基胞嘧啶和 6-甲基腺嘌呤，大肠杆菌 DNA 中也含有一些甲基化碱基。某些噬菌体中存在大量修饰碱基(参见第一章)。

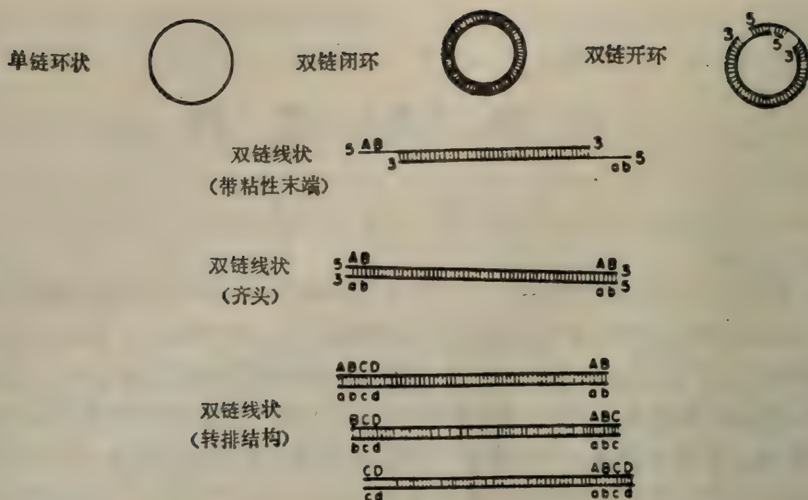


图 2-1 细菌、病毒和细胞器 DNA 分子的不同结构形式
图中 3 为 3′，5 为 5′

以上是有关 DNA 分子的一些概况，以下就 DNA 的一级结构，DNA 顺序组织，DNA 的二级和三级结构分别讨论。

一、DNA 的一级结构

DNA 是双螺旋的大分子。生物信息绝大部分都贮存在 DNA 分子中，这些信息以核苷酸不同的排列顺序编码在 DNA 分子上，核苷酸排列顺序变了，它的生物学含义也就不同了。DNA 的一级结构就是指核苷酸在 DNA 分子中的排列顺序。因此测定 DNA 的碱基排列顺序是分子生物学的基本课题之一。我们首先了解一下怎样测定 DNA 的一级结构，其次将讨论某些特定的序列有些什么生物意义。

尽管测定 DNA 顺序是一个十分重要的课题，然而在不久以前它还是一个令人望而生畏的难题。DNA 顺序测定工作大大迟于蛋白质的顺序测定，也晚于 RNA 的顺序测定。用于蛋白质或 RNA 的顺序测定的战略都是建立在小片段重叠的原理上。也就是把待测分子用专一性的水解酶水解成很小的片段，测定每个小片段的顺序，找出能把这些小片段连起来的重叠片段，然后推断整个分子的顺序。但是这个战略不完全适用于 DNA 顺序测定。存在的问题首先在于没有相应的碱基专一的脱氧核糖核酸酶，因此要获得稳定的小片段十分困难；其次，DNA 分子量一般都很大，获得重叠片段的困难将随着分子量的增加而增加。由于这些固有的困难使得 DNA 顺序测定工作长期以来进展不大。很明显，测定 DNA 顺序的战略如果仍建立在小片段重叠的原理上，那么即使发现了对碱基专一的脱氧核糖核酸酶，测定 DNA 顺序所面临的困难比 RNA 或蛋白质要大得多。看来，DNA 顺序测定工作要有所进展，必须突破原有的顺序测定的战略。七十年代后期这个难关已经克服，主要是从两个方面进行的：首先是限制性内切酶的发现和广泛应用，现在很容易完成含有几千个以上的核苷酸大片段的重叠；其次是建立了一种在凝胶电泳

板上直接而快速地测定 DNA 顺序的方法。在这种新方法中，碱基顺序的测定不是建立在小片段重叠的原则上。

在较为详细地讨论这些突破之前，先简单地回顾一下在此以前探索 DNA 顺序测定的方法，这不仅仅是一个历史的回顾，其中有些技术目前也还有一定的用途。

简单的回顾

七十年代初，DNA 顺序测定还处在探索时期。人们从各个角度摸索测定 DNA 顺序的方法。当时人们的思路还没有跳出小片段重叠的框框，因此所发展起来的 DNA 顺序测定方法工作量都比较大，而且缺乏通用性，只能适用在一些特定的例子中。

例如 RNA 顺序测定方法建立以后，人们就想先把 DNA 转录成 RNA，测定了 RNA 顺序以后，应用互补原理来推测 DNA 的顺序。比较成功的例子是大肠杆菌的乳糖操纵子调控基因这一段 DNA 顺序的测定。但是在多数条件下转录的起始点和终止点是不专一的，因此产生的 RNA 是不均一的，有时比模板还要长，一般 DNA 的二条链都转录出来。如果能获得一段与待测 DNA 互补的引物，转录产物就比较均一。但在很多情况下，选择合适的引物是很困难的，这些都是使用转录方法测定 DNA 顺序的重要障碍。

用一种核糖核苷酸取代 DNA 中的脱氧核糖核苷酸，从而可以获得在特定的碱基位置专一地降解的片段，这在当时是一个很诱人的设想。例如用 DNA 聚合酶在试管中复制待测 DNA，反应中所必须的四种核苷三磷酸底物，其中三种仍然是用脱氧核糖核苷三磷酸，一种如用核糖鸟苷三磷酸代替，在这样的条件下合成的 DNA 在鸟嘌呤的位置都是核糖核酸键，而核糖核酸键可以用碱或核糖核酸酶（RNase）降解，用这种方法可以达到碱基专一性水解的目的。但是在这种条件下，复制是否真实，还是一个有争论的问题，而且在这个方法中需要特定的引物，因此现在已没有人用这种方法测定一个较长的 DNA 片段的顺序了。

用双向电泳同系层析方法分析 20 个核苷酸以下的 DNA 片段具有特殊的优点。一个寡核苷酸片段，其 5′末端（或 3′末端）用放射性同位素标记以后，用外切酶逐个切去非标记末端的核苷酸，这样就得到 5′末端都是带标记的，片段长度只差一个核苷酸的寡核苷酸混合物。将这组寡核苷酸混合物先进行电泳，各个片段基本上按电荷大小分离，然后进行同系层析，使其按不同链长分离。当一个寡核苷酸末端添加一个新的核苷酸时，会影响寡核苷酸的电泳迁移率，影响的程度由添加上去的核苷酸性质所决定，因此可以根据迁移率的变化逐个推断出加上去的是哪一种核苷酸。只是迁移率还受到其他一些因素的影响，这会给判读带来困难。但是由于它的简便性，这个方法在某些情况下还是有用的，例如用于测定限制性内切酶的识别顺序。这个方法也能用于 RNA 小片段的顺序分析。

用二苯胺和甲酸处理，DNA 分子中含嘌呤的糖苷键都断裂，而含嘧啶的键不受影响，这样得到嘧啶连在一起的寡核苷酸称为嘧啶堆或嘧啶群。用双向电泳同系层析可以测定嘧啶堆的组成，借助一些酶的帮助还可以推断嘧啶堆的顺序。这个方法因其简便，也可以用来验证用其他方法测定的顺序的可靠性。某些顺序特别简单的 DNA，例如豚鼠卫星 DNA 顺序就是用这个方法测定的。

测定噬菌体 λDNA 粘性末端顺序的工作可以认为是成功地测定 DNA 顺序的第一个例子。这是在 1971 年前后利用粘性末端本身作为模板引物体系，用 DNA 聚合酶把不

同组合的标记核苷酸参入到新合成 DNA 链内，分析新合成的链中标记核苷酸的种类和数量可以推断原来粘性末端的顺序，用这个方法测定 DNA 顺序的工作量是十分浩大的。λDNA 粘性末端只有十二个核苷酸，而用这个方法测定它的顺序却花费一、二年之久，因此这个方法无法推广。但是利用 DNA 聚合酶在模板引物体系上进行有限的合成，这对以后发展起来的"加减法"测定顺序的方法是有启发的。

在这段时期也有人探索直接测定 DNA 顺序的方法，但是由于沿用测定 RNA 顺序的老路，而又没有类似核糖核酸酶的有碱基专一性的脱氧核糖核酸酶，困难就显得更加突出。1974 年完成含四、五十个核苷酸的 φX174 DNA 片段的顺序测定花费了很大的力量，按这样的效率去完成含有五千多个核苷酸的整个 φX174 DNA 的顺序是十分困难的。

总之，在这一段探索时期，人们只能对一段为数不大的特定 DNA 片段用特定的方法测定其顺序，方法上普遍性不足，而工作量却很大。DNA 顺序测定的突破是从 1975 年 Sanger 发展的"加减法"而开始的。在讨论这个方法之前，还应该先介绍一下限制性内切酶和 DNA 的物理图谱。

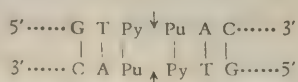
限制性内切酶

微生物中的修饰和限制现象是五十年代初发现的，这是细菌自卫的一种方式。例如从大肠杆菌K株繁殖出来的噬菌体 λ(K) 只能感染大肠杆菌K株，不能感染大肠杆菌B株，反之亦然。这个现象称为"限制"现象。然而被某菌株限制的噬菌体群体中也会有极少个体幸存下来，幸存者的子代可以在原来受限制的菌株中正常繁殖，这个现象称为"修饰"。修饰和限制的性质都是由宿主的基因所决定的，同噬菌体的基因无关。

Arber 和他的合作者对这个现象作了深入研究，并从生物化学角度提出一种假设。他认为细菌内有两种不同功能的酶，一种是核酸内切酶，它能识别并切开 DNA 分子上一个特定的碱基序列，这就是限制性内切酶；另一种酶是甲基化酶，它能识别限制性内切酶所识别的碱基顺序，并把其中某些碱基甲基化，这是修饰酶。被修饰酶甲基化了的 DNA 就不能被限制性内切酶降解了。细菌的 DNA 总是被自身的修饰酶修饰好了，因此不会被自身的限制酶所降解。异源的 DNA 由于没有被修饰过，它一旦侵入细菌就会被限制酶降解，因此这个系统可以保护细菌免受外来 DNA 的入侵。

六十年代末已分别从大肠杆菌K株和B株分离到限制性内切酶，这些酶具有 Arber 所预言的若干特性，如它能降解同限制性内切酶不同源的 DNA，而不降解同源的 DNA。这个酶在 DNA 分子上仅切开为数不多的几个切口。可惜这两个限制性酶都不能在 DNA 分子上产生专一性的切口。

重大的进展是在 1970 年。Smith 及其同事在流感嗜血杆菌 d 株中分离到一个限制性内切酶，并且发现这个酶可以在下列顺序中箭头所指的位置切开，这个酶现在称为 Hind Ⅱ。此处 Py 和 Pu 分别代表嘧啶和嘌呤。



显然，限制性内切酶的这种性质是一种十分奇异的特性。人们也立即认识到这种性

质对 DNA 结构和功能的研究将会带来巨大的好处。从此以后寻找新的限制酶以及把限制性内切酶应用于 DNA 研究的工作广泛地开展起来。

限制性内切酶的命名是用 Nathans 等人所提议的方案。限制性内切酶名称的第一个字母取自获得此内切酶的细菌的属名的第一个字母，用大写，名称的第二、三个字母取自该细菌种名的头二个字母，用小写。如果该细菌还有不同的株系，则另加第四个代表株系的字母或数字，最后是用罗马字大写的数字、代表同一菌株中不同限制性内切酶的编号。如 Hind III 代表从流感噬血杆菌 d 株 (haemophilus influenzae) 中分离到的第三种内切酶。

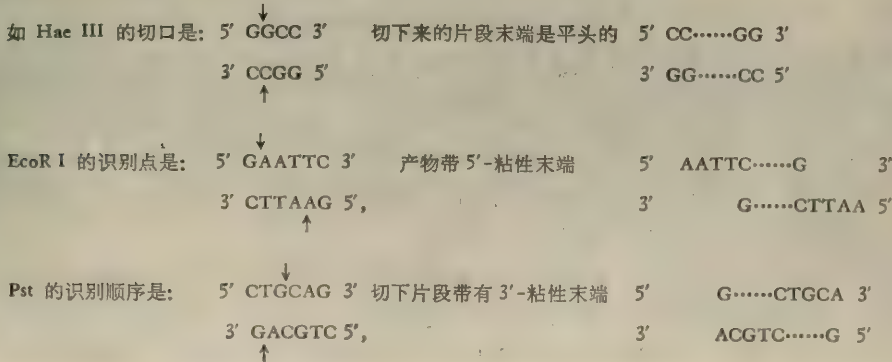
到 1982 年为止已分离到的限制性内切酶有 400 种以上(见附录四)，主要是从细菌和霉菌中得到的。现在也有人报告可以从枝原体和真核细胞中分离到限制性内切酶。已搞清识别顺序的限制性酶约 300 种以上，但其中很多具有相同的识别顺序。具有相同的识别顺序而来自不同菌种的限制性内切酶称为同工异源酶。按识别顺序分类总共约有五十多种识别顺序。很明显作为工具酶，不同专一性的限制性内切酶越多，对 DNA 结构功能的研究的帮助也越大。

用作工具酶的限制性内切酶属于所谓第二类限制性内切酶，它们共同的特点是在 DNA 分子上产生的切口的位置是专一的。此外，这类内切酶的分子量一般不大，通常小于 100,000。镁离子是唯一必需的辅因子。限制性内切酶的识别顺序通常只有 4—6 个碱基对。一个分子量不太大的 DNA 中所包含的内切酶识别点是有限的，因此酶切后产生的片段的大小和数目都是固定的。这些片段用凝胶电泳分离后的电泳图谱对每一种酶而言是具有特征性的，这是目前最常用的鉴定内切酶的种类和活力的基本方法。

从附录四中所列出的限制性内切酶的识别顺序可以看出多数酶的识别顺序是很严格的，少数酶的识别顺序有一定的变动余地。如前面所提到的 Hind II，它的识别顺序中 5′ 端第三个碱基可以是 C 或 T，第四个碱基可以是 A 或 G。大多数识别顺序具有二次旋转对称性，就是说以对称中心为轴。把这个序列转动 180°，仍然得到原来的序列。

从附录四还可以看出，限制性内切酶在 DNA 双链分子上的切点绝大多数位于识别顺序上，只有少数酶的切点在识别顺序之外，而且它们的识别顺序都不具有旋转对称性。如 Hga I 的识别顺序是 GACGC，而切点在这个识别顺序之后 5—10 个碱基。

DNA 分子被限制性内切酶切开之后，它的片段可能有三种形式：一种是 DNA 片段的末端是平齐的，一种是带有 5′粘性末端，一种是带有 3′粘性末端。



正因为有了限制性内切酶，分子量很大的 DNA 才有可能切成一定大小的片段，只有用这些较小的片段，DNA 顺序测定才成为可能。一个 DNA 分子用限制酶酶解后得到很多片段，那么这些片段在 DNA 分子中原来的排列次序如何？这就是下一节物理图谱要解决的问题。

物 理 图 谱

物理图谱就是 DNA 分子的限制性内切酶酶解片段的排列顺序。物理图谱可以用多种方法来测定，其中以部分酶解方法和两种不同专一性限制酶交叉酶解方法最为简便和常用。部分酶解方法是其他各种测定物理图谱方法的基础。

测定物理图谱的第一步是测定 DNA 完全酶解片段的数目和每个片段的分子量。选择合适浓度的琼脂糖凝胶或聚丙烯酰胺凝胶作为载体进行电泳，通常能把绝大多数酶解片段分开。这些片段的分子量能从泳动的距离来判断，因为在合适的条件下泳动的距离同 DNA 分子量的对数成反比。如果 DNA 片段是放射性 ^{32}P 全标记的，那么就可以从每个片段的放射性占整个分子的放射性量的百分比推算出分子量。用电子显微镜方法可以测定每个 DNA 片段的长度，从而计算出分子量，但没有电泳方法方便。

测定了完全酶解片段的数目和分子量之后，接着把 DNA 用限制性内切酶进行部分酶解，测定每个部分酶解片段的分子量。从部分酶解片段的分子量就可以直接推测哪几个完全酶解片段是相邻的。例如有一个 φX174（复制型）Hind II 酶解片段，其分子大小是 1030 碱基对（bp），这个片段用 Alu I 可以切成 A，B，C，D 四个片段，分子大小分别是 750bp，150bp，80bp 和 50bp。为了测定这四个片段的排列顺序，可以用 Alu I 部分酶解，在部分酶解片段中得到分子大小分别为 810bp，230bp 130bp 等片段。从这些数据立即可以推断 Alu I 酶解的四个片段的排列次序为 150-80-50-750(见图 2-2)。当然也可以把每个酶解片段从凝胶上洗下用同样的酶完全酶解，这样得到的酶解片段应是相邻的片段。

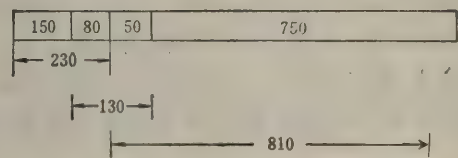


图 2-2 用部分酶解方法测定物理图谱

用两种限制性内切酶交叉酶解也是测定物理图谱常用的方法。当由于某种原因，例如高级结构的影响某些部分酶解片段产率特别低时，这就需要第二种酶来帮忙。交叉酶解不仅可以得到两种或多种限制性内切酶的物理图谱，而且可以相互校对，使结果更加可靠。一般先建立产生片段比较少的限制性内切酶的物理图谱，然后在这个基础上去建立产生片段比较多的限制性内切酶的物理图谱。

如前所述，部分酶解的方法是建立物理图谱的基本手段，但是部分酶解片段的数目大大多于完全酶解片段的数目，这给分离带来很多困难。现在有一种改进的部分酶解方法，先把 DNA 分子的一端用放射性同位素标记，部分酶解后电泳分离，那末只有带同位素

的片段出现在放射自显影图上，谱线的数目相当于完全酶解的片段的数目，相邻谱线的碱基对数目之差就是两个相邻点之间这段 DNA 片段的大小。例如某个组蛋白基因片段共5900 bp，末端标记之后，用 Hpa II 部分酶解后得五个片段，其分子大小分别是 5900bp，4210 bp，3260bp，2950bp，1020bp，显然，最大的部分酶解片段和其次一个片段核苷酸数之差是最靠右边的片段的分子大小，其余以此类推，就可以推得全酶解五个片段的排列次序是：1020-1930-310-930-1690（见图 2-3）。

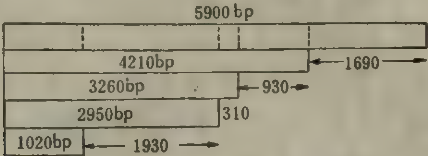


图 2-3　用末端标记法测定物理图谱

测定物理图谱的方法还有很多，如引物脉冲标记法、双向杂交法、电子显微镜法、基因互补法、基因删除突变株和野生株比较的方法、外切酶逐步酶解法等，这里不一一详述了。测定了某一 DNA 分子的物理图谱之后，就可以着手测定每一片段的核苷酸排列顺序了。

测定 DNA 顺序的方法

前面已经讨论过，用于 RNA 或蛋白质顺序测定的基本战略是小片段的重叠。而这种战略用于 DNA 顺序测定困难就很大，几乎是行不通的。1975 年 Sanger 突破了这个框框，提出了一种新的战略，他不是直接测定 DNA 片段中碱基的组成和顺序，而是以测定片段的分子长度来推测碱基的顺序，用不同的处理方法来决定碱基的性质。其后发展起来的各种 DNA 顺序测定的新方法都是以这个原理为基础的。近来这个原理也用到 RNA 顺序测定中，发展了很多新的 RNA 顺序测定的新方法(见第三章)。因此这个原理的提出在核酸的顺序测定中是有划时代意义的。

这个方法的原理并不复杂。假定有一个 DNA 片段，它的 5′末端用放射性同位素标记好了，这是片段长度的参考位置。又假定这个 DNA 片段中有二个腺嘌呤，以 5′末端为起点，一个在第六位，一个在第十位。如果我们设计一种方法，能使与腺嘌呤相邻的磷酸二酯键断开，而与其他碱基相邻的链不受影响，那么将可以得到二个片段，一个是六核苷酸，另一个是十核苷酸，这二个片段可以用电泳方法分开，并决定它们的长度。反过来，如果测得片段的长度，我们可以推测，以 5′末端为参考位置，第六位和第十位由腺嘌呤占据着，其他三种碱基的位置可以用类似的方法推测。显而易见，这个新方法能否成功取决于两点，首先是要有一种方法能把仅仅相差一个核苷酸的同系核苷酸片段都分开，这一点现在是办得到的，例如在 40 厘米长的聚丙烯酰胺凝胶上电泳就可以把 100—200 核苷酸都分开，分辨率达到一个核苷酸的水平。其次要有一种手段能制备四种末端碱基专一的片段。1975 年 Sanger 设计的“加减法”首先满足了这个要求。从此以后发展了很多制备片段的方法。我们所讲新的 DNA（或 RNA）顺序测定方法，其原理都是如此，不同的方法区别就在于制备片段的方法不同。按制备片段的方法不同可以把这些方法分成两大

类：一类是用 DNA 聚合酶，简称酶法；另一类是用化学降解，简称化学法。

酶法　从链长决定碱基位置的新的顺序测定方法首先是由 Sanger 于 1975 年设计的"加减法"实现的。以单链 DNA 为模板，加一个合适的引物（一般用限制性内切酶的酶解片段），再加入四种脱氧核糖核苷三磷酸（dNTP），其中一种是同位素标记的。DNA 聚合酶就从引物开始合成一条与模板互补的 DNA 链。对于 DNA 顺序分析最理想的情况是，使合成反应进行的程度尽可能是随机的，就是使合成产物中各种长度的片段都存在。然后从反应混合物中除去四种 dNTP，反应产物分成二部分，一部分用在加法系统，一部分用在减法系统。用于加法系统的产物再分成四小分，每一小分中仅仅将四种 dNTP 之中的一种加入反应体系，比如仅仅加入 dATP。由于所用 DNA 聚合酶具有 3′→5′的外切酶的活性（见第四章），而反应体系中缺少另外三种必要的 dNTP，合成产物就从 3′→5′方向降解。然而由于存在 dATP，显然遇到腺嘌呤的位置降解反应就停止了。因此所有的片段都是以 A 为结尾的。同理，可以分别制备以 C、G 或 T 结尾的另外三组片段。四组片段在同一块凝胶板的不同样品槽中同时电泳，从放射自显影图上就可以推断碱基顺序，因为从 A 样品槽走出的放射性区带代表 A 在片段中的位置，同理也可以定出 C、G 和 T 的位置。因此根据同一凝胶电泳上的四组区带就可以推断碱基的顺序。按理只用一个加法系统就足以推断 DNA 的碱基顺序了。但是由于技术上的原因，只用一种方法有时不能得到完全正确的结果，因此还设计了一种减法系统。在减法系统中，也是将上述酶促产物分成四小分，每一小分中只加入三种 dNTP，比如缺少 dATP。在这个反应体系中，DNA 聚合酶能够把片段继续合成下去，但是在遇到应该是 dATP 参人的位置

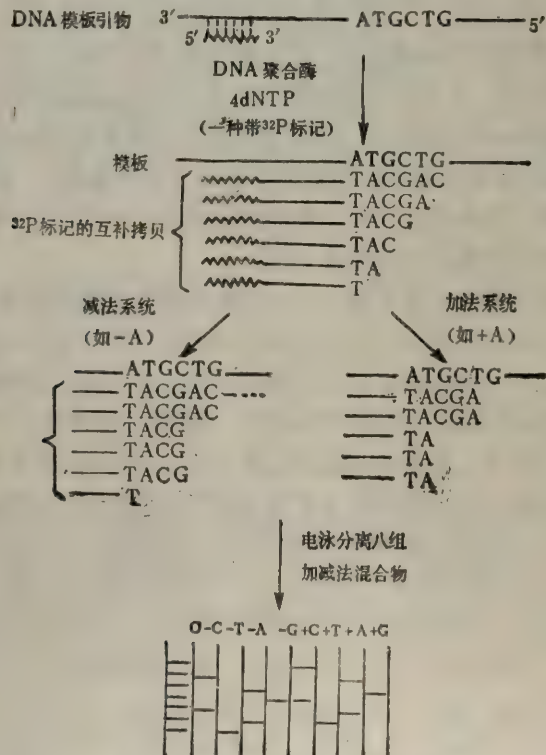


图 2-4 "加减法"测定 DNA 顺序的原理

时，合成反应就停下来了。这就可以得到一组都是以A前一个位置为结尾的片段，电泳后同样可以定出A的位置。

加减法是第一个成功地把片段长度和碱基位置联系起来的新的顺序测定方法，φX 174 DNA 全部顺序就是用这个方法测定的。在 DNA 顺序测定方法的发展过程中，加减法的历史性位置是不可低估的。但由于技术上的问题，这个方法的实际效果还不够理想。可能由于碱基的特异性和 DNA 的高级结构的影响，聚合反应并非如希望的那样不同步，反应速度有差异，有些片段可能多些，另外一些片段可能少些，有时可能导致漏读和重读，有时图谱上还会出现假谱线，当同样的碱基联排时图谱上有时只出现一条带。因此用这个方法测定的 DNA 片段可能有五十分之一的误差。

对加减法的一个最成功的改进是"末端终止法"，其基本反应也是利用 DNA 聚合酶的酶促反应。在反应混合物中有模板、引物、酶和四种 dNTP。除了四种正常的 dNTP 之外，另外再加入一定比例的 2′，3′双脱氧核糖核苷三磷酸，比如 ddATP (图 2-5)。因此在

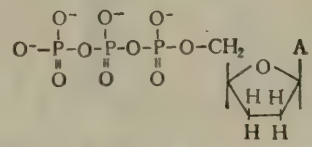


图 2-5 2′、3′ 双脱氧核糖腺苷三磷酸的结构

合成过程中遇到应该 dATP 参入的位置就有二种可能的情况发生。如果是 dATP 参入则链可以继续延伸，如果是 ddATP 参入合成反应就不能进行了，因为 ddATP 的 3′位置没有羟基。因此，可以得到一系列不同长度的以 ddA 为结尾的一组片段。同理，也可以得到以 G、T 或C为结尾的片段，电泳之后，就可以直接"读出"碱基顺序，这是一个十分方便和精确的方法。当然如果原始的 DNA 是双链，而不是单链，那么分离单链 DNA 片段是困难的，而且制备合适的引物也是一个棘手的问题。1980 年以来发展起来的把DNA重组到单链噬菌体 M13 中，增殖之后，可以获得大量的单链 DNA (酶法正需要单链 DNA 作为模板进行顺序测定)，而且由于组建的 M13 载体含有通用的引物结合位置，因此可以利用同一种引物，这样简化了单链 DNA 分离和引物制备的步骤，大大推动了末端终止法的发展。以这个方法为主，已于 1982 年完成了 λ 噬菌体 DNA全顺序 (48502bp) 的测定，目前正向更大的 DNA 进军。

1982 年洪国藩改进了上述双脱氧与 M13 克隆系统相结合的方法。由于原有方法是随机的，不可避免地出现测定的多次重复，而这种重复在测定的后期更为严重。改进后的方法叫做"DNA 顺序的连续测定法"(见示意图 2-6)。连续法的关键是将克隆的外源 DNA 的一端固定起来，而另一端则用没有专一性的 DNase I 将其逐渐缩短，然后通过"单规扫描"找到长度约有 200 个核苷酸差异的克隆；实验设计使得缩短的一端发生在共同引物区，这样，对这些克隆的顺序的逐个测定，就实现了对长链 DNA 的连续测定。用这个方法完成了 2300bp 的顺序。

化学法 Gilbert 及其同事 Maxam 利用某些化学试剂能专一地修饰碱基，使相应的糖苷键变得不稳定，这样也可以得到以特定碱基结尾的四组片段 (图 2-7)。他们所选用的化学试剂一种是硫酸二甲酯，一种是肼。硫酸二甲酯可以使鸟嘌呤的 N-7 位和腺嘌

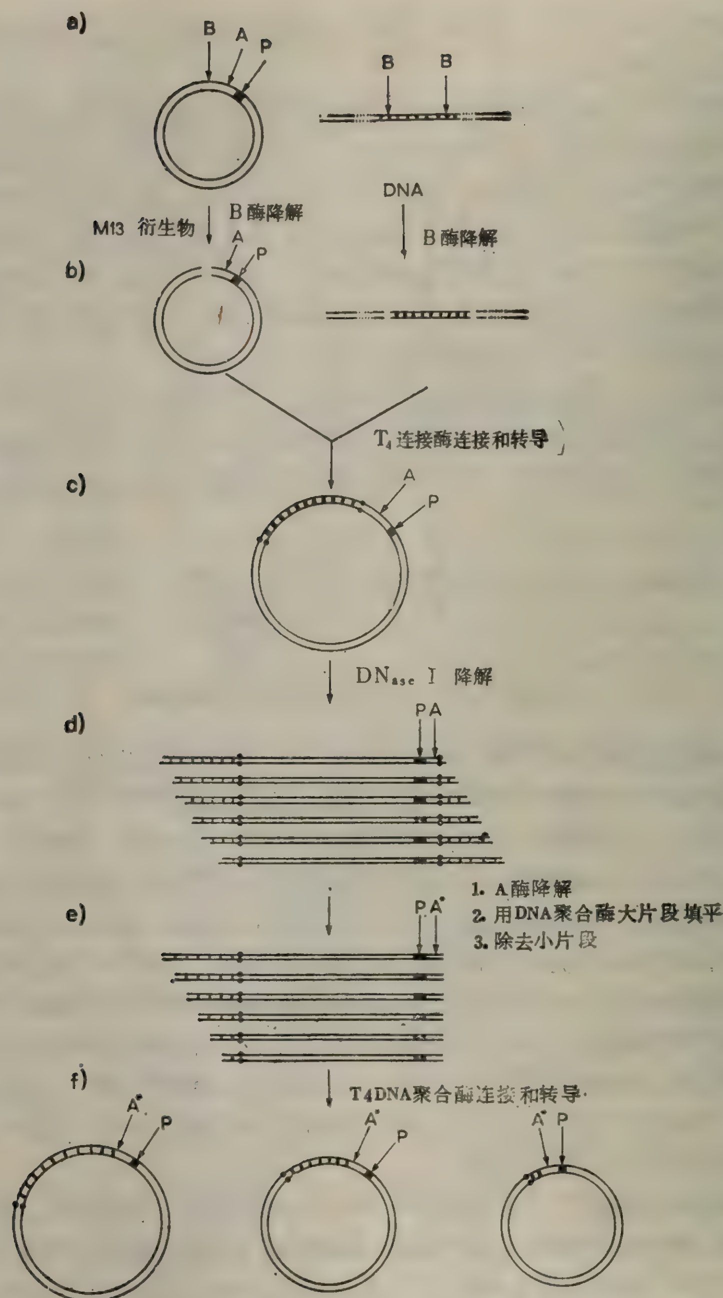


图 2-6 连续法测定 DNA 顺序示意图。 A和B为二个不同的限制性内切酶

呤的 N-3 位甲基化，被甲基化的嘌呤的糖苷键是不稳定的，在中性 pH 加热就会使甲基化的嘌呤脱落，当脱嘌呤的 DNA 在碱性溶液加热时，磷酸二酯键就会断开，这样一组片段电泳之后在放射自显影图上有一深一浅二组谱线，深的是由于G降解而产生的，浅的是由于A降解而产生的。G之所以比A深是因为G甲基化的速度比A快五倍。用硫酸二甲酯甲基化的 DNA 片段，如果先用稀酸处理，脱嘌呤反应主要是在A的位置上进行，因为在这种条件下甲基化的A比甲基化的G更不稳定，然后再用碱断开磷酸二酯键，电泳图谱

上深的谱线代表 A，浅的谱线代表 G。对于嘧啶可以选用水合肼部分肼解，然后用哌啶催化 β 消除反应，使磷酸基团掉下，此条件对 C 和 T 没有选择性，当肼解时存在 2M NaCl，.

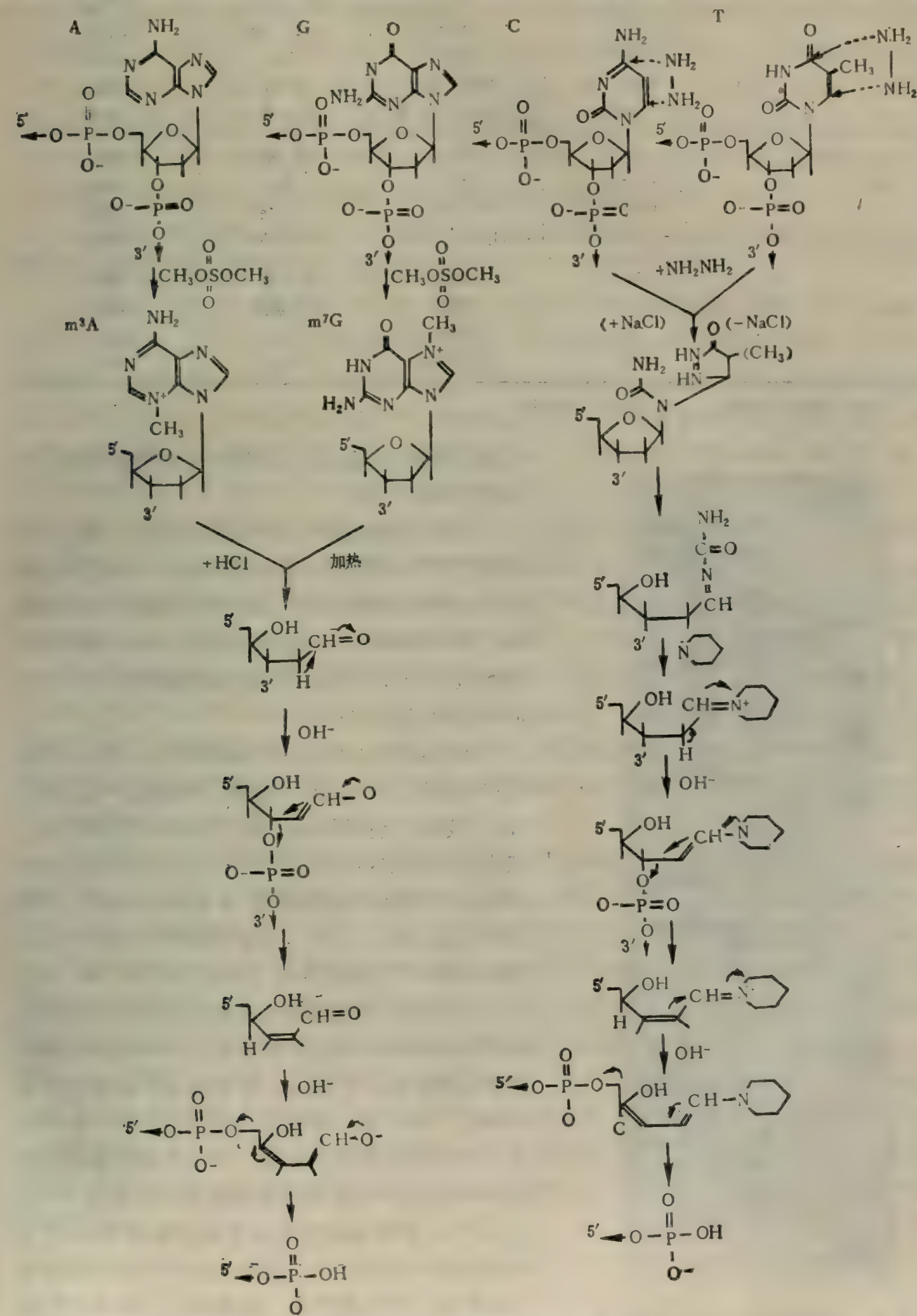


图 2-7　DNA 化学降解的反应历程

表 2-2　DNA 顺序分析的化学降解反应

组　别		断　裂	碱基修饰	碱基消除	断链条件
I 组	1	G>A	DMS*	pH7 加热	NaOH
	2	A>G	DMS	酸	NaOH
	3	C+T	NH_2NH_2	哌　啶	哌啶
	4	C	NH_2NH_2 + 盐	哌　啶	哌啶
II 组	5	G	DMS	哌　啶	哌啶
	6	G + A	酸**	酸	哌啶
	7	C + T	NH_2NH_2	哌　啶	哌啶
	8	C	NH_2NH_2 + 盐	哌　啶	哌啶
III 组	9	A>C	NaOH	哌　啶	哌啶
	10	G>A	DMS	pH7 加热	哌啶
	11	G	甲烯蓝	哌　啶	哌啶
	12	T	四氧化娥	哌　啶	哌啶

* DMS 为硫酸二甲酯。　　** pH2 甲酸吡啶。

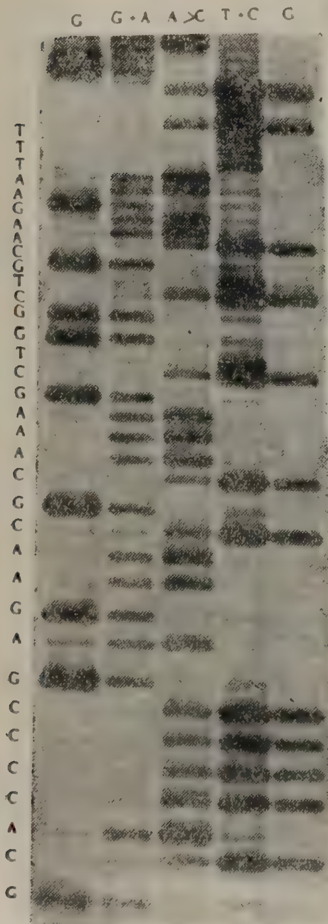


图 2-8　用化学法测定 DNA 顺序的放射自显影图(共采用五个反应条件)

抑制了肼同T的反应，这样产生的谱线相当于C的位置(图 2-7)。

在对二种碱基专一的化学反应中，酸处理通过质子化作用使嘌呤糖苷键不稳定，碱处理优先打开腺嘌呤和胞嘧啶环，最后均可被哌啶所消除，使 DNA 链在G和A(酸处理)或A和C(碱处理)处断链。此外，甲烯蓝，四氧化娥也曾被用于 DNA 顺序分析(见表 2-2)。

如果单链 DNA 的一端被标记，用上述方法可产生以同一种碱基（或同二种碱基）结尾的片段，只要能区别四种碱基，便可直接在凝胶电泳上读出核苷酸排列顺序。表 2-2 列出的三组化学降解反应可被使用，其中以第 II 组效果较好。若在第 II 组再加上 A > C 反应(见第 III 组的 9)，共五个反应条件，能得到更为满意的结果 (图 2-8)。

化学降解的精确程度在理论上应达到100％，但在实际测定中并非如此理想。如果用厚的聚丙烯酰胺凝胶板 (1.5mm)，没有其他手段核对并且仅测定一次的话，精确率可能只有 85％；如果从二个方向测定则可提高到 99％；如果用薄的凝胶板(0.3mm)，一次测定的精确度就可以达到 99％。

DNA 顺序测定方法主要是前面讨论的化学降解法和末端终止法，也有人设计一些其他的专一性断裂 DNA 的方法，但基本上还是模仿化学降解和末端终止法。事实上，除了这二种方法以

外,其他的方法很少有人采用。化学降解法和末端终止法比较,应该说各有利弊。对单链 DNA 而言,末端终止法是特别合适的,而化学降解法由于它的条件比较简单,很多实验室也很乐意采用。近年来由于将末端终止法与 M13 克隆系统相结合已经把 DNA 顺序分析推向一个新阶段。由于方法上的突破,DNA 顺序测定已成为分子生物学中发展最快的分枝之一。报道 DNA 顺序测定的文献应接不暇。第一个完整的 DNA 分子的已知结构是 φX174DNA (5386, 1977 年),接着测定了 SV40 DNA (5226, 1977 年)、G4 DNA (5577,1978 年) 和 fd DNA (6408,1978 年)。1981 年报道了人和牛线粒体 DNA (16569bp 和 16338bp) 的全结构。1982 年发表的 λ 噬菌体 DNA (48502bp)。

在测定大的 DNA 过程中,必须使用电子计算机储存、处理和比较已经获得的顺序信息,这些工作单纯依靠人力已很难胜任了。

几年前,人们还千方百计设法把 DNA 转录成 RNA,用测定 RNA 的方法来反推 DNA 的顺序。现在,DNA 顺序测定的方法已达到这样的水平,就是把 RNA 变成 DNA 来测定。已完成的一些最大的 RNA 的顺序如卵白蛋白的 m RNA、16S rRNA 等都是经过反转录成 DNA 后测定顺序的。另一方面,把核酸分子特异性地降解,然后用测定片段长度来确定碱基位置也已用于 RNA 顺序测定上,这方面近几年有很多发展(见第三章“RNA 结构”)。总的来说,RNA 顺序测定进展还不如 DNA。

由于 DNA 顺序测定方法已十分简便,因此只要有可能,就把蛋白质顺序测定转化成 DNA 的顺序测定。例如猴病毒 SV40 全部顺序测定已完成,只要分离到 SV40 编码的蛋白质,测定其N端开始的几个氨基酸,就可以很容易地按氨基酸密码表推断出蛋白质的全结构。又如噬菌体 φX174 的全部顺序已知后,按照一些规则可以从这个顺序预言 φX174 可能还编码一些至今我们未知的蛋白质。事实也是如此,φX174K 基因蛋白就是在这个预言指导下新找到的 φX174 编码的蛋白质。DNA 顺序测定还可以核对过去测定的蛋白质的顺序是否正确。例如大肠杆菌 i 基因的顺序测定完成后,比较该 DNA 顺序和阻遏蛋白的氨基酸顺序,发现两者有所不同。后来知道,这个差别是由于在氨基酸测定中漏掉了一段肽段。

如果对 DNA 顺序测定作一个乐观的估计,可以设想现在所有的 DNA 顺序原则上都能测定。主要是工作量的问题。我们可以设想,先把一个生物体的全部 DNA 做成基因库,每个克隆里面包含了几千个到几万个碱基对长度的 DNA 片段,按目前的水平,每一个这样长短的 DNA 片段的顺序可以在合理的时间内把它测定完毕。下一步是如何确定每个克隆所包含的 DNA 片段在整个染色体上的排列。可能有很多途径来解决这个问题,例如制备同一种 DNA 用两种不同限制性内切酶酶解后制备两套基因库,分别测定他们的顺序之后就可以找到必要的重叠片段,从而把基因库中的 DNA 片段排列成一个完整的染色体。另一种方法是把每个基因库所含的基因表达出来,同遗传学得到的基因图对照,也可决定很多基因库的顺序,象大肠杆菌 DNA 的顺序在现有水平上是能够解决的,以上的估计可能乐观一些,问题在于即使已知全部顺序,要理解它们的功能将困难得多。

二、DNA 顺序组织

DNA 分子中不同的区域有不同的功能，有些区域是编码蛋白质的结构基因，有些包含了复制、转录或翻译等调控信号。具有不同功能的 DNA 区域在整个分子中的分布情况就是 DNA 顺序组织所要讨论的主题。在原核生物中结构基因占了整个 DNA 的大部分，而在真核生物中，可能只占一小部分。人们相信非结构基因部分包含了调节基因表达的信号。它们的分布情况对基因的调控是直接有关的。研究 DNA 顺序的组织是结构功能研究的一部分。从广义上讲，DNA 一级结构也是 DNA 顺序组织的一部分，一级结构可以看作为最精细的 DNA 顺序组织。当然从另一个角度而言，染色体上的基因图也属于 DNA 顺序的组织，因此 DNA 顺序组织也称为基因的组织。在这一节中我们主要从 DNA 顺序的角度讨论它们的分布特点。

测定 DNA 顺序的组织可以用不同的方法，主要有：

(1) 测定 DNA 核苷酸顺序。近年来核苷酸顺序的快速测定方法已经解决，这为了解 DNA 顺序的组织打下了良好的基础。但是测定 DNA 核苷酸顺序并不等于了解 DNA 顺序的组织，还要与生化的和遗传的方法相配合，了解这一段 DNA 的生物功能，这样才能对顺序组织有比较清晰的理解。

(2) 核酸杂交。用已知功能的 DNA 或 RNA 片段同 DNA 杂交，再配合电子显微镜的观察，是测定 DNA 顺序组织十分有效的手段。特别是近几年来，利用重组 DNA 技术能获得大量特定的探针，这样就为研究基因的数目和分布提供了有利的条件。

(3) 复性动力学方法。测定变性 DNA 的复性动力学的过程可以推测同源 DNA 片段的数目和分布情况。十多年前，这个方法在探索真核生物的基因组织中作出了很好的开端，现在仍然是一个十分有用的手段。

(4) 遗传学方法。利用基因突变的重组、交换和互补等方法可以测定基因的数目和位置，还能测定基因的精细结构。在微生物中，遗传学的方法显出特殊的优点，因为微生物的突变容易获得和检出。

以下我们通过几个例子来简要地说明原核生物中和真核生物中 DNA 顺序组织的某些特点。由于生物的复杂性，特别在真核生物中，因此以下讨论的某些 DNA 顺序组织的特征有可能仅仅是一种特殊的现象，而不是一种普遍的规律。

原核生物 DNA 顺序组织

噬菌体

噬菌体 φX174 和 G4

大肠杆菌噬菌体 φX174 和 G4 都是小球形病毒，包含单链环状 DNA。这二个病毒在血清学上有交叉反应，然而在遗传学上两者不能有效地互补和重组。从电子显微镜观察，这二种 DNA 的异源杂交双链有 39% 碱基是误配的，二者的物理图谱的差别更加证实了这一点。现在二个 DNA 的全部顺序已测定完毕，对这二个噬菌体的遗传学分析和生化研究也十分详尽，因此可以对这两个分子进行详细的比较。令人惊奇的是两者虽然顺序上的差别可达将近 40%，而在 DNA 顺序组织上又那么相似(图 2-9)。从 DNA 顺

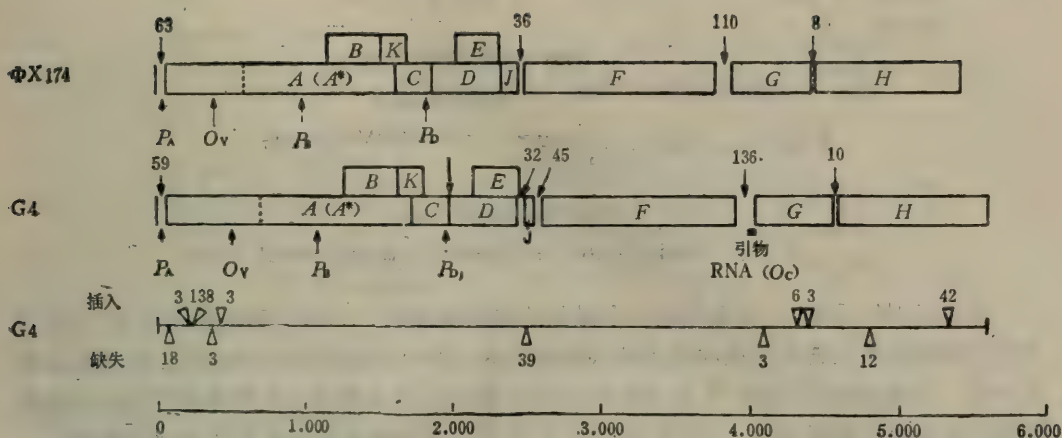


图 2-9 噬菌体 ϕX174 和 G4 的基因图

第一、第二行分别是 ϕX174 和 G4 的基因图；第三行表示相对于 ϕX174 的 G4 编码区中插入和缺失的核苷酸的数目和位置；第四行是核苷酸数目尺度。P_A、P_B 和 P_D 为三个启动子的位点，O_V 和 O_C 分别表示噬菌体 DNA 链和互补 DNA 链合成的起始部位。

序组织的角度比较噬菌体 ϕX174 和 G4，以下三个特点是值得注意的。

(1) 虽然 ϕX174 和 G4 都编码了 11 个蛋白质，但是只转录成三个 mRNA。每个 mRNA 包含了多个顺反子，而且功能上有关的顺反子通常串联在一个 mRNA 分子上。例如从启动子 P_D 开始转录的 mRNA 包含了基因 D-(E)-J-F-G-H。其中基因 J，F，G，H都是编码噬菌体的外壳蛋白的。D蛋白同噬菌体的装配有关，E蛋白是裂解细菌的蛋白。这些编码在一个 mRNA 分子中的蛋白质在生理功能上都是密切相关的。这种 DNA 顺序组织的方式可能是使这些基因协同表达的一种调控方式。功能上相关的基因组织在一起，并转录在同一个 mRNA 分子中，这种现象在原核生物中是很普遍的。在真核生物中，如果说不能绝对排除这种现象，至少也是很稀有的。

(2) ϕX174 和 G4 DNA 分子的绝大部分是用来编码蛋白质用的。只有一小部分是不翻译出来的。不翻译出来的部分在 ϕX174 DNA 中占 217/5386，在 G4 DNA 中占 282/5577，也就是说不到 5%。染色体的绝大部分用来编码蛋白质，这也是原核生物的 DNA 顺序组织的一个特点。不翻译出来的 DNA 区域，或称间隔区，经常包含了一些控制基因表达的 DNA 顺序。如基因H和A之间的间隔区包含了 RNA 聚合酶的结合点，转录的终止信号、核糖体结合位点等控制表达的部分。对噬菌体 ϕX174 而言，所有的间隔区(除了基因G和H之间的间隔之外)都可以形成一个稳定的发夹结构。据推测控制信号可能同这类二级结构有关。

(3) 对噬菌体 ϕX174 和 G4 而言，在 DNA 顺序组织上最大的特点是基因重叠。其他一些病毒，如猴病毒 SV40 DNA 中也发现有基因重叠的现象，至今还只是在病毒 DNA 中发现这种现象。一般而言，这一类病毒 DNA 分子都不大，但又必须装入相当多的基因，可能这种压力导致病毒 DNA 在进化过程重叠起来。从图 2-9 可以看到，基因 B 和基因 K 重叠在基因 A 之内，基因 E 重叠在基因 D 之内。这些重叠在一起的基因使用的码组是不同的，因此虽然是同样的 DNA 顺序，而转译出来的是不同的蛋白质。图 2-10 列举

了部分核苷酸顺序，它们用在三个码组之中，含意各不相同。

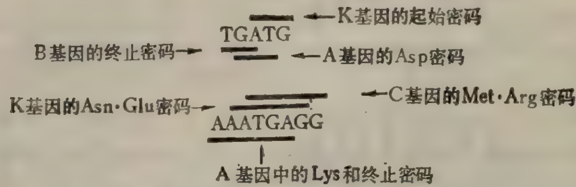


图 2-10　在重叠基因中同一核苷酸顺序包含着多种生物信息

由于基因的重叠，在重叠部位一个碱基的突变将影响二个蛋白质甚至三个蛋白质，因此我们可能猜想在进化过程中位于基因重叠区的 DNA 顺序的保守性将比非重叠区的大一些。实验比较了 φX174 和 G4 的 DNA 顺序的变化频率，发现重叠区的变化频率仅仅比非重叠区稍低一点。(22% 对 33%)，因此基因的重叠并不是进化的严重阻碍。

噬菌体 λ　　噬菌体 λ 是感染大肠杆菌的温和性噬菌体。λ DNA 入侵宿主以后，可以有二条发育途径，一条是溶菌途径，另一条是溶原途径。所谓溶菌途径就是入侵的 λ DNA 在宿主内增殖，指导合成噬菌体的 m RNA，并翻译出噬菌体的蛋白质，装配成成熟的噬菌体，使宿主细胞裂解。所谓溶原途径就是入侵的 λ DNA 整合到大肠杆菌的染色体中，λ DNA 基因的表达受到阻遏，λ DNA 作为宿主 DNA 的一部分与宿主 DNA 一起复制。在某些条件下，比如紫外光照射之后，可以诱导 λ DNA 从宿主 DNA 上切下，独立增殖，发育成成熟的噬菌体，使宿主细胞破裂。λ 噬菌体的这二条发育途径可以看作一种最原始的分化现象。

噬菌体 λ DNA 顺序组织具有一般原核生物所具有的共同特点，如大部分 DNA 是用来编码蛋白质的，功能上相关的基因组织在一起，并且一起转录成一个 m RNA，这样的组织方式同调控的关系就不详述了，这里仅讨论 cI 基因两边二个操纵基因的结构(图 2-11, 2-12)。

λ DNA 入侵细菌之后，首先利用宿主的生物合成系统分别从 P_L 和 P_R 转录翻译成二个蛋白质，N 和 cro 蛋白。这二个 m RNA 在转录终止信号 t_{L1} 和 t_{R1} 位置转录结束。这二个终止信号是受 ρ(宿主转录终止因子)控制的。噬菌体编码的蛋白 N 有抗 ρ 因子的能力，因此蛋白 N 的合成就使转录可以延伸到 t_{R1} 和 t_{L1} 之外，进一步合成蛋白 cII，O，P，cIII 等基因产物。蛋白 O 和 P 是引发 λ DNA 复制的蛋白质，cII 和 cIII 是同噬菌体进入溶原途径有关的蛋白质。什么因素决定 λ 噬菌体的发育途径这个问题还没有解决。如果是进入溶原途径，由于蛋白 cII 和 cIII 是正调节因子，在它们的帮助下，使得从 P_{re} 开始的转录能够起动。转录一直进行到 cI 基因，产生蛋白 cI。蛋白 cI 是一个阻遏蛋白，它结合到操纵基因 O_R 和 O_L 上，阻止了从启动子 P_R 和 P_L 开始的转录。这时 cII 和 cIII 的转录也被阻遏蛋白所阻遏，cI 的转录从 P_{rm} 开始，这时阻遏状态就确立了。在阻遏状态，λ DNA 仅仅合成阻遏蛋白 cI，蛋白 cI 结合在操纵基因 O_R 和 O_L，阻止了所有 λ 基因的表达。图 2-11 列出 λ 噬菌体的部分基因排列图。特别列出与转录调控有关的部分，与复制调控有关部分将在第四章第五节讨论。

噬菌体 λ 中和转录调控有关的操纵基因 O_R 和 O_L 分别由 80 个左右碱基构成，结构上十分相似，两者又可大致分成三个区域(见图 2-12)。下面我们着重分析 O_R 中的细节

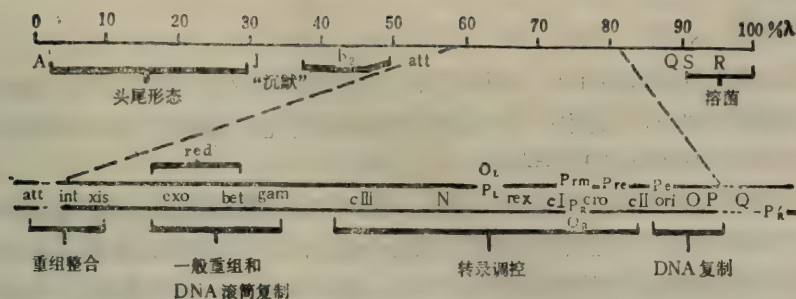


图 2-11 噬菌体λ的部分基因图

图中 CI 应为 cI，CII 为 cII，CIII 为 cIII

问题。在这个操纵基因区域有三个阻遏蛋白结合点，分别称为 O_{R1}，O_{R2} 和 O_{R3}。每一个结合点包含 17 个核苷酸，二个结合点之间有若干个核苷酸间隔。而 O_{R1}、O_{R2} 和 O_{R3} 的核苷酸顺序十分相似，仅稍有差异，构成一个小的重复家族。这些顺序的相似而不相同对调控有微妙的影响。因为这些顺序都很相似，所以都能结合阻遏蛋白 cI。但又因为这些顺序有所区别，它们的结合能力稍有不同。O_{R1} 结合阻遏蛋白 cI 的能力最强，O_{R2} 次之，O_{R3} 最弱。当环境中阻遏蛋白的浓度不高时，它首先结合到 O_{R1} 阻止 P_R 的右向转录，但是不影响从 P_{rm} 的左向转录。因为这二个启动基因都在 O_{R1} 和 O_{R3} 之间，这样可以保证 λDNA 处于阻遏状态，而阻遏蛋白 cI 可以继续合成。当 cI 蛋白合成多了，它就

N O_{L1} O_{L2} O_{L3}（O_L） cI O_{R3} O_{R2} O_{R1}（O_R） cro

O_{R1}	T	A	T	C	A	C	C	G	C	C	A	G	A	G	G	T	A
O_{R2}	T	A	A	C	A	C	C	G	T	G	C	G	T	G	T	T	G
O_{R3}	T	A	T	C	A	C	C	G	C	A	A	G	G	G	A	T	A
O_{L1}	T	A	T	C	A	C	C	G	C	C	A	G	T	G	G	T	A
O_{L2}	C	A	A	C	A	C	C	G	C	C	A	G	A	G	A	T	A
O_{L3}	T	A	T	C	A	C	C	G	C	A	G	A	T	G	G	T	T
	T_5	A_6	T_4	C_6	A_6	C_6	C_6	G_6	C_5	C_3	A_4	G_5	T_5	G_6	G_3	T_6	A_4
	C_1		A_2						T_1	A_2	C_1	A_1	A_2		A_2		G_1
										G_1	G_1		G_1		T_1		T_1

图 2-12 噬菌体λ操纵基因 O_R 和 O_L 的顺序

不仅同 O_{R1} 结合，也可以同 O_{R2}、O_{R3} 结合了，这就阻遏了 cI 基因的自我表达。因此 cI 蛋白的浓度是可以自我调节的，它不受细菌本身生产情况的影响，而维持 λDNA 处于阻遏状态。以后我们将看到真核细胞中有很多重复顺序家族，对这些顺序的功能我们还了解很少。我们能否认为 O_{R1}、O_{R2} 和 O_{R3} 这种 DNA 顺序组织形式是真核生物中重复顺序的原始形式？如果低等的噬菌体λ的重复顺序有如此巧妙的调节功能，那么真核生物中的重复顺序是否有更加巧妙的调控功能？这就需要进一步用实验证据来回答了。

大肠杆菌

大肠杆菌染色体是由 4×10^6 核苷酸对组成的双链环状 DNA，对大肠杆菌的基因图已有大量的详细分析，其中一些片段的 DNA 顺序也已测定。从 DNA 顺序组织的角度来说以下几个方面值得着重讨论。

(1) 同前面提到的几种噬菌体一样，大肠杆菌的很多基因也是以操纵子的形式组织起来的。操纵子包括功能上相关的几个蛋白质基因和一个共同的调控区域，如乳糖操纵子包括了三个编码酶蛋白的结构基因（Z, Y, A）和一个相连的调控区，即启动基因(启动子)、操纵基因、CAP 结合区等。这些功能上相关的结构基因转录在一个 mRNA 分子上(图 2-13)。

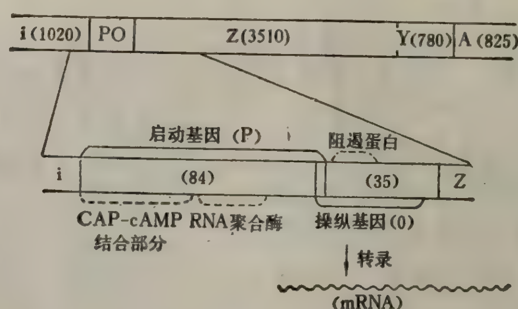


图 2-13 乳糖操纵子的顺序组织
括弧内为核苷酸数目，调控区的 DNA 顺序见第五章图 5-12

另一方面，也并不是所有功能有关的基因都组织在一起，例如同大肠杆菌复制有关的十多个基因散布在整个染色体的各个部分，而并没有组织在一起成为一个所谓复制操纵子。我们知道 DNA 复制的调控是一个远比乳糖操纵子的调控更为复杂的过程，因此似乎可以这样说，如果一组基因同一个复杂的生理功能有关，那么这些基因不组织在一起可能更有利于利用复杂的调控方式。

(2) 以 RNA 为最终产物的基因，它们的拷贝数往往不止一个。大肠杆菌中编码核糖体 RNA (rRNA) 的基因有 7 个拷贝，每个拷贝都是以 16S—23S—5S 的顺序串联在一起（详见第五章），尽管每个拷贝都转录出同样的核糖体 RNA，但是这些 RNA 基因前后的核苷酸顺序每个拷贝都是不同的。这些差别对核糖体 RNA 转录的调控有些什么影响还很不清楚。

大肠杆菌转移核糖核酸 (tRNA) 基因总共有 60 个拷贝，主要分布见第五章图 5-34。大肠杆菌 tRNA 基因并不串联在一起构成一个大的转录单位，而是分成若干小的转录单位，一般每个转录单位含二、三个 tRNA 基因，受同一个启动子控制，因此一个转录单位内的 tRNA 的产量是一样的，而不同转录单位之间的 tRNA 产量可能不一样。现在还不知道不同的转录单位是否包含能共同协调的调控顺序。某些 tRNA 基因被组织在 rRNA 操纵子之中。

(3) 位于转录控制区域的 DNA 顺序的组织形式是多种多样的，不同的组织形式与不同的生物功能有明显的关系。

最早被阐明转录调控区域的 DNA 顺序组织是乳糖操纵子(见图 2-13)。调控区由二

个基因（其中有三个结合部位）组成（图 2-13），三个结合部位从左至右排列是 CAP-cAMP 结合部位、RNA 聚合酶结合部位和阻遏蛋白结合部位，紧接在后是 z 结构基因。当转录开始时，RNA 聚合酶首先同启动子（P）结合，然后右向滑动，到达一定部位开始转录。当阻遏蛋白与操纵基因（O）结合，就会在空间上妨碍 RNA 聚合酶滑向转录开始部位，因而不能产生与乳糖代谢有关的酶。但如加入诱导物（如乳糖或其类似物），可与阻遏蛋白形成复合物而使阻遏蛋白失活，结果基因表达开放，转录出一条多顺反子 mRNA 链，继而翻译出三个相关的酶（图 2-14）。RNA 聚合酶结合部位的 5′上游是 CAP-cAMP 结合部位。CAP（降解代谢活化蛋白）是 cAMP 的受体蛋白。当细胞内葡萄糖供应不足时，cAMP 浓度比较高，因此结合到乳糖操纵子上的 CAP-cAMP 复合物也比较多，CAP-cAMP 是一个正调因子，这时转录增加；但当葡萄糖供应充分时，cAMP 浓度下降，阻碍了许多降解代谢基因包括乳糖操纵子的转录。这是一种很合理的调控方式，因为葡萄糖供应充分时，细菌没有必要通过降解其他种类的糖来获得能量。当用乳糖和葡萄糖同时培养细菌时，细菌首先利用葡萄糖的碳源，只有在葡萄糖耗尽后，经过一个潜伏期才开放乳糖操纵子以产生降解乳糖的酶，由乳糖供给碳源（图 2-15）。CAP-cAMP 同 DNA 结合以后，可能有利于 RNA 聚合酶同启动子结合，从而发挥正调作用，它与作为负调因子的阻遏蛋白表现为相反的调节作用。

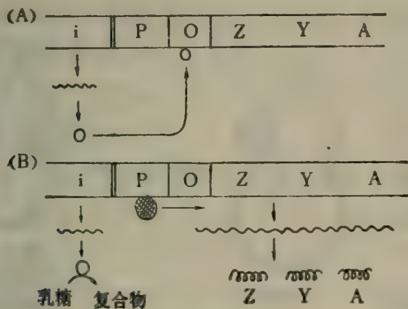


图 2-14 乳糖操纵子的阻遏(A)与诱导(B)
~~~，mRNA； O，阻遏蛋白； ⊛，RNA 聚合酶；Z，β-半乳糖苷酶；Y，半乳糖苷通透酶；A，半乳糖苷乙酰化酶。

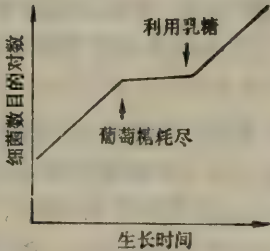


图 2-15 用葡萄糖和乳糖为碳源培养大肠杆菌

半乳糖操纵子的调控区域的 DNA 顺序组织同乳糖操纵子稍有不同，半乳糖操纵子中 DNA 顺序组织的次序从左到右是操纵基因、启动子，CAP-cAMP 结合部位位于操纵基因而不是启动子中。如前所述，CAP-cAMP 是一个正调因子，当它结合到半乳糖操纵子时就加快了转录。这时转录从+1 开始。但是当阻遏蛋白结合到操纵基因时，使转录水平大大降低而且转录从−5 位开始。但不完全破坏 RNA 聚合酶与启动子的相互作用，即并不完全关闭半乳糖操纵子 mRNA 的转录。相反，乳糖操纵子结合了阻遏蛋白之后，转录几乎完全停止了。这两个操纵子对阻遏蛋白的反应性不同与它们的 DNA 顺序组织有关，乳糖操纵子的操纵基因在启动子右边，一旦结合了阻遏蛋白使 RNA 聚合酶完全不能通过，而在半乳糖操纵子中，操纵基因在启动子的左边，结合阻遏蛋白之后，并不会妨碍 RNA 聚合酶的通过，它的效应可能在于阻断了 CAP-cAMP 的正调效应。图
~~~

2-16 是半乳糖操纵子调控区的 DNA 顺序组织。

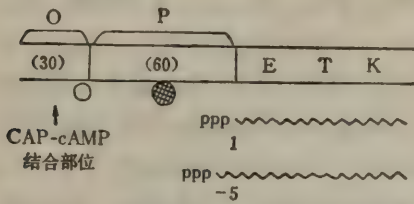


图 2-16 半乳糖操纵子的 DNA 顺序组织

符号同图 2-14。调控区位于左侧，核苷酸数常用负号表示，P 在 1 至 -60 处，O 位于 -60 至 -90。

通过以上二个例子我们可以看到 DNA 顺序的组织对代谢调控方式是有很重要影响的，而这种不同的调控方式可能与不同的生理效应有关。半乳糖是大肠杆菌糖蛋白的必要的组成成分，即使半乳糖不用作能源，少量的代谢半乳糖的酶仍然是必需的；而乳糖仅被用作细胞的能源，如果细胞环境中有其他能源时，如上所述，细菌就没有必要合成有关乳糖代谢的酶。

在操纵子的调控部位中，除了上述的启动子、操纵基因和 CAP-cAMP 结合点等单元之外，七十年代还发现了一个新的调控单元，称为衰减基因（衰减子，attenuator）。衰减子都位于启动子之后，也都存在于同氨基酸合成代谢有关的操纵子中。它们表现出共同的生理现象，当体内某种氨基酸丰富时，转录进行到衰减子就停止了。因为细胞没有必要去合成已经富裕的氨基酸；如果体内某种氨基酸缺乏时，转录可以通过衰减子，一直转录到结构基因，并翻译出相应的蛋白质（酶），合成相应的氨基酸。关于色氨酸操纵子的顺序组织见第五章图 5-17。1981 年测定了该操纵子中 7,000 多个核苷酸的排列顺序，已能从一级结构水平上阐述衰减子的调控方式（详见第五章，第 159 页）。

(4) 很久以来人们就为细菌抗生素抗性的迅速传播所困惑：常用的抗生素对不少病菌已经无能为力，各种新的抗生素用于临床以后不久，就会发现有些病菌对它们也产生了抗药性。经多方面的探索发现，其原因之一就是抗药基因可以借助于转座子由一种细菌的质粒上转移到另一种细菌的质粒上去，如图 2-17 所示。

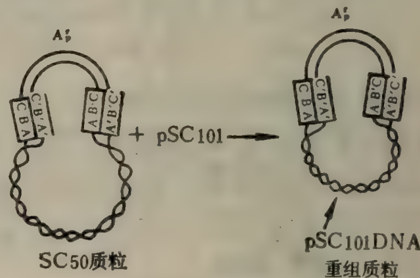
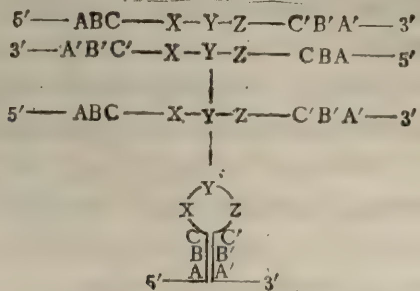


图 2-17 抗氨苄青霉素（A_p^r）DNA 片段从 pSC50 质粒转位到 pSC101 质粒上

转座子（transposon）或转位因子（transposable element）的 DNA 通常有两段反向的重复顺序，该反向重复顺序的长度可由几对核苷酸到 2,000 对核苷酸。当将转座子的双链 DNA 拆开成单链后，在适当的条件下退火，转座子的单链 DNA 可形成环状发夹结构：



转座子可以把彼此无关的 DNA 片段连接在一起使人猜想其中必定有某种能专一地识别反向重复顺序的酶在起作用。果然，实验证明转座子 Tn3 中带有转座酶基因和一种编码阻遏物的基因。在转座酶基因区域内如发生突变就会影响转座子的转位，而在阻遏物基因区的突变可以增加 Tn3 在不同质粒间转移的效率。

虽然转座子可以将它自身及连接于其中的其他基因一同插入到受体 DNA 分子的许多位点上，但这种插入并不是随机的。受体分子的某些部位特别容易被转座子插入，这些被称作"热点"的顺序与转座子两端的反向重复顺序非常相似。

转座子不仅可以在质粒之间转移抗药基因，也可以转移其他性状(抗汞化合物，合成细菌毒素，代谢各种碳水化合物……)的基因。转座子的受体分子可以是质粒 DNA，也可以是细菌染色体或病毒 DNA。因此转座子可以实现基因片段的来回转移和染色体上遗传信息的重排。

以上极其简单地讨论了原核生物中 DNA 顺序组织的某些特点。可以看到，即使在低等生物中，DNA 顺序组织也是很复杂的。以下我们将讨论真核生物的 DNA 顺序组织。很显然要总结出真核生物 DNA 顺序组织的规律就更加困难了。

真核生物 DNA 顺序组织

真核生物 DNA 比原核生物 DNA 大得多，DNA 顺序组织也相应地复杂得多。有关真核 DNA 的顺序组织的认识最早来自对 DNA 复性动力学的研究，这样得到的有关 DNA 顺序组织的知识是粗线条的。近几年来，借助于 DNA 重组、杂交、核苷酸顺序测定等技术，对真核 DNA 顺序组织有了比较深入的了解，但其深度远远比不上对原核 DNA 的认识。

从 DNA 复性动力学观察，真核 DNA 可以分成三个部分。一部分是能很快复性的 DNA，这种 DNA 是由几种不太长的核苷酸顺序重复几百万次构成的，这种顺序称为高度重复顺序。如卫星 DNA 就是一种高度重复顺序。另一种是中度重复顺序，他们的重复次数从几十次到几千次之间。如 rRNA 基因、tRNA 基因、某些蛋白质的基因(如组蛋白，肌动蛋白，角蛋白等)。第三类是单拷贝基因，它们在整个基因组织中只出现一次或少数几次，如珠蛋白基因，卵白蛋白基因，丝心蛋白基因等。

有些高度重复顺序 DNA 的浮力密度同主体 DNA 有区别，在密度梯度超离心时，它独立于主体 DNA 而自成一个小峰，卫星 DNA 的名称就是由此而来的。高度重复顺序 DNA 可能是由极其相似的含有 2—10 个碱基对的顺序片段重复多次而构成。有些卫星 DNA 的基础顺序已测定，如豚鼠卫星 DNA 是以 CCCTAA 为基础顺序，其他的顺序都是从这个顺序经过突变而衍生出来的。后来发现卫星 DNA 的基础顺序可能要长得多。例如非洲绿猴 DNA 用限制性内切酶 Hind III 酶解，总共 7% DNA 都是以 172 个碱基对的基础顺序重复了几百万次而构成。高度重复顺序一般位于异染色质上，可能不转录成 RNA，它的功能还不清楚。近年来发现在高度重复顺序中也间隔了可以转录的顺序。

用不同长度的 DNA 作探针，发现真核 DNA 中单拷贝顺序和重复顺序是间隔排列的。这里讲的重复顺序是指从复性动力学看复性比较快的、功能上未知的一类 DNA 片段。这种间隔排列可以分成二类，一种称为爪蟾型，一种称为果蝇型。绝大多数真核 DNA

都属于爪蟾型。它的组织形式是约 300bp 长度的重复顺序。中间间隔了大约 800—1,000bp 长度的单拷贝顺序。重复顺序约占整个基因组的 25%。也发现少量长度约 2,000—6,000bp 长度的重复顺序，约占 5%。在果蝇型中，重复顺序长达 3,500—5,000bp，间隔的单拷贝顺序的长度大于 13,000bp。近来也发现有介于二者之间的组织类型，如鸡的 DNA 顺序组织就是如此。

这些重复顺序按顺序相似程度可以分成不同的家族，每个家族的大小可以不同，但都包含一些成员。重复顺序大多是可以转录的，不同的家族转录的频率也不同。有人指出，重复顺序的转录有明显的组织特异性。在多核糖体内没有发现这些重复顺序的转录产物，可能它们仅局限于细胞核内。

复性动力学是研究 DNA 顺序组织一个很好的手段，但是用这个方法得到的是统计学上平均的结果，是一种粗线条的概貌，要得到比较精细的结果必须结合其他的手段。以下就不同类型的基因的组织作些简单的讨论。

最终产物是 RNA 的基因

这里指的是编码核糖体中 RNA 的基因和编码 tRNA 的基因。这些基因的产物都是同蛋白质的生物合成有关的 RNA，在细胞内的含量比较高，都属于中度重复顺序。

rDNA　在真核 DNA 中，rDNA 一般重复程度达几百次。如酵母中 rDNA 重复 140 次，果蝇中 130—250 次，爪蟾中 400—600 次，人是 300 次左右。真核生物的核糖体包含四种 RNA，即 28S，18S，5.8S 和 5S RNA。在低等真核生物中，这四种 RNA 的基因都串联在一起。

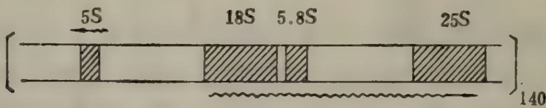


图 2-18　酵母 rDNA 的顺序组织

图 2-18 是酵母的 rDNA 的顺序组织，图中带斜线的部分代表 rRNA 基因，余为基因间的间隔，带箭头的波纹线表示转录的方向和长度。从图中我们可以看到，18S、5.8S 和 25S 组成一个转录单位，5S 构成一个独立转录单位，两个转录单位的方向是相反的。同时，基因和基因之间有间隔区，间隔区可分为两类：一类是不转录 RNA 的，如 5S 和 18S 之间的间隔区；另一类是可以转录在原始 RNA 产物中，如 18S 和 5.8S、5.8S 和 25S 之间的间隔区，但不出现在成熟的 RNA 分子中，这些间隔区在 RNA 成熟过程被修剪掉了。

在比较高等的真核 DNA 中，5S 基因不同其他 rRNA 基因串联在一起，因此 rDNA 一般只包含 26S，5.8S，18S 三个基因串联在一起的一段 DNA。如家蚕的 rDNA 的组织如图 2-19 所示。

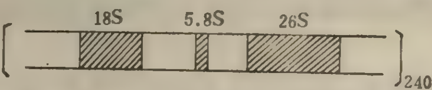


图 2-19　家蚕 rRNA 的顺序组织

每个重复单元的大小是 6.9×10^6，其中 6×10^6 是转录成 rRNA 的前体，1×10^6 的顺序

不转录成 RNA。家蚕 rDNA 的间隔区组成和长度都是均一的，这种例子并不多见，在大多数生物中 rDNA 的间隔区在长度上和组成上都有些变化。

有些 rDNA，如果蝇的 28S 基因被插入顺序（intervening sequence）分开。间隔区是指基因之间的不出现在成熟产物中的顺序，而插入顺序则是指基因（结构基因）内部的不出现在成熟产物中的顺序；在 RNA 成熟即加工过程中，间隔区只要删除即可，而插入顺序非但要删去，而且还要把断开的 RNA 拼接起来才能保持基因产物的完整性（关于加工细节见第五章）。

5S DNA　上面已经提到了，在比较高等的真核生物中 5S DNA 独立组成一个转录单位，不与其他 rRNA 基因串联在一起。不同的生物中 5S DNA 的组织也很不同。例如人的 5S DNA 仅仅集中在染色体的一个部位上，果蝇的 5S DNA 位于染色体 II 的右臂，而爪蟾的 5S 基因却散布在 18 个染色体中的 15 个上面。

果蝇的 5S 基因组织比较简单。果蝇染色体共有 160 个 5S 基因的拷贝串联在一起，前 80 个拷贝和后 80 个拷贝在功能上可能独立起作用而不相互干扰。5S基因中还没有发现插入顺序。每个拷贝之间的间隔区的长度和组成稍有变化，特别有趣的是 5S 基因的间隔区不转录成 RNA，原始的转录产物就是成熟的 5S RNA。

在爪蟾中有 9,000—24,000 个 5S 基因，爪蟾 5S 基因包括四种不同的 5S RNA，差别最大的两种 5S RNA 有七个碱基不同。在爪蟾卵中，合成所有这四种 5S RNA；在肾细胞中只合成一种。肾细胞中所表达的 5S 基因可能只占总的 5S 基因的 3%。分析了爪蟾 5S 基因的核苷酸顺序以后，发现有一个特殊的现象，就是除了 5S 基因之外，紧接着一个不转录的 5S 基因，称为伪 5S 基因。它比正常的 5S 基因短 20 个碱基对。这种伪基因起什么作用目前还一无所知。

5S RNA 是不经过前体直接从基因上转录出来的产物，如前面很多例子所表明的，一般启动基因都是在基因的 5′ 上游。然而用重组 DNA 的方法制备了一系列不同位置碱基缺失的 5S 基因，然后再进行体外转录，发现了一个十分奇特的现象，即 5S 基因的启动基因不在它的 5′端，而在基因本身的中间。这在基因组织上是一个新的现象，位于基因本身中间的启动基因如何控制基因的转录也是一个有趣的问题。

tRNA 基因　tRNA 是从 tRNA 前体剪切得来。tRNA 前体平均长度约 140 个核苷酸。在爪蟾中有 40 种不同的 tRNA，平均每种 tRNA 基因有 200 个拷贝，这 200个拷贝串联在一起构成一个重复单位。果蝇中有 75 个 tRNA 基因拷贝。在酵母中约有 320—400 个 tRNA 基因。假定酵母的 tRNA 种类也是 40 种，那么每一种 tRNA 基因约有 8—10 个拷贝。

近年来分析了很多 tRNA 基因的碱基顺序，发现有二个特点，第一，尽管编码同样的 tRNA，如都是编码酪氨酸 tRNA 的基因，也尽管转录出的 tRNA 完全一样，但基因前后的碱基顺序每个成员都不一样。第二，有些 tRNA 基因有插入顺序，而有些则没有，同样是酪氨酸 tRNA 基因，有的拷贝中有 18 个核苷酸的插入顺序，而有的是 19 个核苷酸的插入顺序。现已分离纯化了一个剪接酶，它能正确地切除 tRNA 中的插入顺序，并把两个片段连接成完整的 tRNA 分子。剪接酶的种属特异性似乎不是很专一。酵母的带插入顺序的 tRNA 前体也能在爪蟾卵母细胞中正确地剪接。

以上是从 DNA 顺序组织角度简单地讨论了最终产物为 RNA 的基因组织的情况，

较为详细的内容以及转录和加工问题参阅第五章。从以上的介绍我们可以看到，这些RNA的基因组织是极其复杂的，没有一个共同的规律可循。在不同的生物体内这些基因的组织是不同的，即使在同一生物体内，编码同一产物的基因的组织也是多样的，即使是同一种产物，有时也有细微的差别。这些组织上的差别可能具有功能上的意义，但是我们对此知之甚少。

最终产物是蛋白质的基因

最终产物为蛋白质的基因有几个特点，分述如下：

（1）蛋白质基因的数目可能是单拷贝的，也可能是多拷贝的。多拷贝基因的产物，如组蛋白，肌动蛋白，角蛋白等，一般都属于需求量比较大的蛋白质。但是需求量大的蛋白质的基因不一定都是多拷贝的，例如珠蛋白基因，卵白蛋白基因都是单拷贝基因，尽管在红血球中绝大部分蛋白质是血红蛋白，在鸡卵白中，绝大部分蛋白质是卵白蛋白。

（2）在功能上相关的基因可以串联在一起，也可能相隔很远。以珠蛋白为例。血红蛋白包含二条 α 珠蛋白链，二条 β 珠蛋白链，在人细胞中，α 链和 β 链的基因不是串联在一起，而且不在同一个染色体上，前者位于染色体 16，后者位于染色体 11。α 链除了能和 β 链构成正常成人的 $\alpha_2\beta_2$ 血红蛋白之外，还能同 γ 珠蛋白和 δ 珠蛋白构成 $\alpha_2\gamma_2$ 和 $\alpha_2\delta_2$ 血红蛋白。在胎儿血液中主要是 $\alpha_2\gamma_2$ 血红蛋白，一旦出生后，血液中血红蛋白的类型就逐步变成 $\alpha_2\beta_2$ 型，以及少量 $\alpha_2\delta_2$ 型。β, δ, γ（包括 $^A\gamma$ 和 $^G\gamma$）珠蛋白都属于 β 类型。β 类型的珠蛋白基因都串联在一起，中间以间隔区隔开。每种类型的珠蛋白都是分别转录的（图 2-20）。

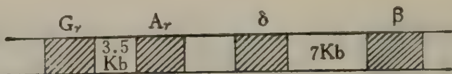


图 2-20　人 β 类型珠蛋白基因的排列（Kb 代表千碱基对）（图中 Gr 应为 $^G\gamma$, Ar 应为 $^A\gamma$）

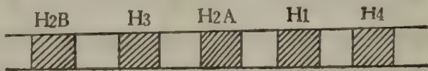


图 2-21　组蛋白基因的排列

图 2-21 是组蛋白基因的组织，细胞核内组蛋白一共有五种，分别称为 H1、H2A、H2B、H3 和 H4，这五种蛋白质的基因串联在一起，构成一个重复单位。在海胆中，这个重复单位长约 5,400 bp，在整个基因组中重复几百次。虽然过去有人认为组蛋白基因可能转录在一个 mRNA 中，然后分别翻译出不同的组蛋白，即认为组蛋白基因是多顺反子的。但最近的实验结果表明五个组蛋白基因是分别转录出各自的 mRNA，就是说组蛋白基因也是单顺反子结构。迄今为止在真核生物中还没有发现一个蛋白质基因是多顺反子结构的，这是真核生物和原核生物蛋白质基因结构的一个很重要的区别。

虽然现在普遍认为真核细胞中，蛋白质基因组织中都是单顺反子结构，也就是说一个基因对应一个 mRNA，一个 mRNA 对应一条肽链，但是这并不排除一个肽链通过后加工可以生成几种不同性质的蛋白质或多肽。例如促肾上腺激素-β-脂肪酸释放因子前体的 mRNA（它的 cDNA 核苷酸顺序已完成，共 1091 核苷酸）能翻译成一条蛋白链。这

个前体蛋白通过后加工能生成促肾上腺素、β-脂肪酸释放因子、促黑激素、β-内非肽等不同的激素，而且从前体的顺序，还能推测可能存在的但目前尚未发现的激素。图 2-22 就表示促肾上腺激素-β-脂肪酸释放因子前体通过后加工可以产生很多种激素。

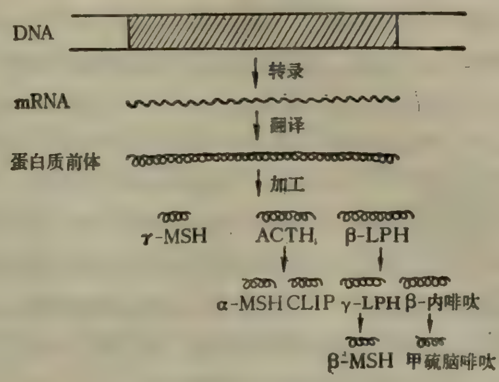


图 2-22 促肾上腺激素-β-脂肪酸释放因子前体及后加工产物
ACTH，促肾上腺皮质激素； MSH，促黑激素； LPH，脂肪酸释放激素； CLIP，促皮质激素样中间肽。

总之对真核生物而言，功能上相关的基因可能串联在一起，也可能相隔很远，但是不论哪种情况，蛋白质基因都是单顺反子结构。

(3) 很多蛋白质基因包含插入顺序

上一节已提到，酵母 tDNA 和果蝇 rDNA 有插入顺序。许多目前研究得比较多的蛋白质的基因都发现有插入顺序(但组蛋白基因没有插入顺序，这一点已从核苷酸顺序测定予以最后肯定)。由于基因上有了插入顺序，使得编码一个蛋白质的 DNA 顺序在基

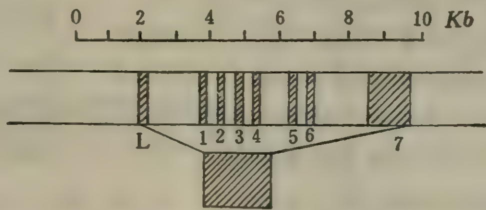


图 2-23 鸡卵白蛋白基因中的插入顺序
带斜线方框为编码蛋白质的结构基因，即外显子，外显子之间为插入顺序，即内含子。图中有七个内含子和七个外显子加一个前导顺序 L。

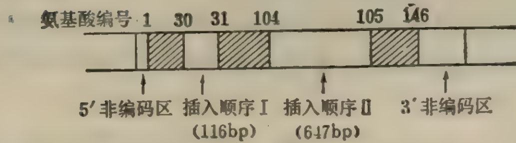


图 2-24 小鼠 β 珠蛋白基因的插入顺序
β 珠蛋白由 146 氨基酸组成，被三个隔开的外显子所编码，基因中除插入顺序外，还转录出两端的非编码区。

因上是不连续的。已有几十种蛋白质基因含有插入顺序，如 α、β、δ 珠蛋白，免疫球蛋白的重链和轻链，卵白蛋白(见图 2-23)，溶菌酶，SV40，腺病毒等。图 2-24 是 β 珠蛋白基因插入顺序的分布情况。插入顺序有一专有名称，称为内含子，编码蛋白质顺序的那部分 DNA 称为外显子(也见第五章第 165 页)。

插入顺序的生物功能还很不清楚。在产生卵白蛋白和不产生卵白蛋白的细胞之间，卵白蛋白基因的插入顺序没有区别，因此插入顺序和转录的效率没有关系。在分化细胞和生殖细胞之间，卵白蛋白基因的插入顺序也没有区别，因此在分化过程中插入顺序没有变化。但是在不同的个体中插入顺序有细微的差异。有些实验结果说明，插入顺序同 mRNA 转录后的命运密切有关；也有人认为插入顺序是对生物进化十分有利的基因组织方式，因为有了这种方式，基因很容易通过重组而产生带有新功能的蛋白质。免疫球蛋白重链的插入顺序的排列方式支持这个假说，因为免疫球蛋白的每个外显子所编码的肽段实际上是一个独立的功能单元。

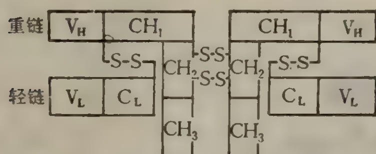


图 2-25　免疫球蛋白肽链的基本结构

图 2-25 是免疫球蛋白的肽链的基本结构。免疫球蛋白由二条重链和二条轻链构成，轻链和重链都可分成二部分，一部分是恒定区，这个区域的氨基酸顺序在不同的抗体中是不变的；另一部分是可变区，可变区的氨基酸顺序随抗体专一性而变化。重链的恒定区又可分成三个部分（CH_1、CH_2、CH_3），每个区域约由 110 个氨基酸组成。CH_1 区域同轻链的 C_L（同 V_H 和 V_L 一起)构成 γ 球蛋白的 F_{ab} 部分，二条重链的 CH_2 和 CH_3 组成 F_c 部分。F_{ab} 和 F_c 由一个绞接区连接。对免疫球蛋白的功能进行深入研究后发现，这些区域各自都有自己特有的功能。V_H V_L 配对，这部分是负责同抗原结合的；CH_1 是轻链的附着；配对的 CH_2 用于补体固定；配对的 CH_3 负责同细胞表面相互作用。还有一个N端信号肽，同球蛋白的分泌有关。分析免疫球蛋白基因的核苷酸顺序后发现一个很有意义的事实，即编码每个功能部分的顺序都被一个插入顺序分开。图 2-26 就示意了免疫球蛋白的轻链和重链的基因结构。从图中可以很清楚地看到每个外显子都编码一个有一定功能的肽链。

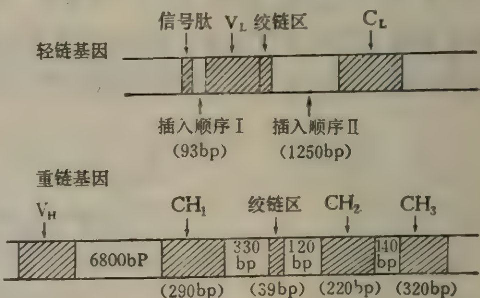


图 2-26　免疫球蛋白重链和轻链基因的结构

β 珠蛋白有三个外显子(见图 2-24)，虽然每个外显子的功能没有象免疫球蛋白区分得那么明显，但是可以发现同血红素结合的氨基酸大多处于中间的那个外显子中。有的实验指出，一个基因中的内含子就是另一个基因的外显子，表明这段 DNA 序列也有编码功能。近来发现，在酵母线粒体中有一个内含子同时存在着起始密码子和终止密码子，能编码一个有 235 个氨基酸的蛋白质，表明所谓内含子其本身可能是一个独立的结构基因。

总结以上例子有人提出，每个外显子实际上是来自一个原始蛋白质，内含子也可能具有生物功能。外显子和内含子镶嵌的方式可能是蛋白质进化的遗迹，而且也是进一步进化的原料。可能正因为有了插入顺序这种基因组织，大大加速了真核生物的进化。

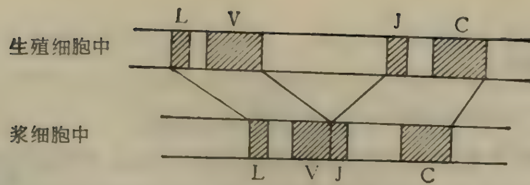


图 2-27　在生殖细胞和浆细胞中 λ_1 基因组织的比较

(4) 在一个生物中，DNA 顺序组织可能是可变的。一个例子是免疫球蛋白 λ_1 链。λ_1 链的基因在浆细胞中，V_L（可变区）和 J（绞链区）是连在一起的，而在胚胎细胞中这二个区域是分开很远的，如图 2-27 所示，在分化过程中通过基因重组 V_L 和 J 区域才连接在一起。这种基因重组的方式是产生多种免疫球蛋白的机制之一。除了在免疫反应过程中基因有重组现象外，在真核生物中也观察到类似原核生物的基因转移。五十年代早期人们就发现，某种可转移的基因决定了玉米粒的色泽和形状。后来又发现果蝇的眼睛的颜色取决于某个可转移的多拷贝基因，因此某些 DNA 片段可以在染色体转位的现象是肯定的。

现在还不知道这种现象是局限于某些生物某些基因的特例，或是在其分化过程中也会产生 DNA 顺序组织的变化。从已有的实验结果似乎说明，在细胞分化过程中，DNA 顺序组织的变化即使存在，也不是一种普遍的现象。

(5) 在真核生物中，可能存在一种原核生物中不存在的基因表达的调控单元，称为"增效子"(enhancer)，猴病毒 40 (SV40) DNA 中存在二段顺序和方向完全一样的 72 碱基对的片段，它对 SV40 早期基因的表达是必须的，把这一 DNA 片段和其他基因(比如 β 珠蛋白基因)连在一起转化细胞以后，测定珠蛋白基因的表达效率，实验结果表明这 72bp 片段大大促进了珠蛋白表达，而且这种效应同 72bp 片段同珠蛋白基因之间的相对距离和方向都没有关系，但和细胞的种类有关。具有这种效应的 DNA 片段称为"增效子"。不但 SV40 具有"增效子"，其他一些肿瘤病毒中也有类似的"增效子"，但碱基顺序差异很大。除病毒外，某些真核基因内也具有这样类似的调控单元。

总而言之，真核生物 DNA 顺序组织是很复杂的，没有一个简单的规律可循。我们对它的认识还处在初步阶段。至于顺序组织和功能的关系更是知之不多，有待于进一步研究。

三、DNA 的二级结构

DNA 双螺旋

DNA 的水溶液是很粘稠的，加热又能使溶液粘度大大降低，这说明 DNA 分子具有一定结构。DNA 分子应该具有怎样的结构才能使它的溶液有这种流体力学性质呢？四十年代 Pauling 提出了蛋白质的 α 螺旋模型，同时人们已经知道 DNA 是遗传信息的载体，那么 DNA 应该具有怎样的结构才能成为基因载体呢？1953 年 Watson 和 Crick 二人提出了 DNA 双螺旋模型，这个模型不仅解释了当时所知道的 DNA 的一切理化性质，而且将结构与功能联系起来，大大推动了分子生物学的发展，其历史功绩怎样评价也不过

分。

双螺旋提出的根据

当时提出 DNA 双螺旋结构的根据主要有三方面。

第一，Astbury 很早就用X射线衍射方法研究 DNA 纤维的结构，后来 Wilkins 等人得到更加精美的 DNA 纤维的X射线衍射图，这些衍射图提示 DNA 分子可能具有螺旋结构。一些强的衍射点表明 DNA 分子中有 3.4 Å和 34 Å的周期性的结构。

其次，Chargaff 分析了大量不同来源的 DNA 的碱基组成，发现不象原来假定的那样，DNA 是由等量的 A、G、C、T 组成。相反，不同来源的 DNA，其碱基比例是不同的，但是有一个共同的规律，即胸腺嘧啶的摩尔含量等于腺嘌呤的摩尔含量，胞嘧啶的摩尔含量总是等于鸟嘌呤的摩尔含量，即 [A] = [T] 和 [G] = [C]。

第三，根据四种碱基的物化数据，碱基可以分成大小二种，嘌呤比嘧啶大些，如果A-T配对或 G-C 配对，二个碱基对的几何大小十分相近；同时这两种碱基配对在化学上是合理的，因为从嘌呤和嘧啶的氨基和酮基以及它们的键长键角来看，A-T 之间以及 G-C 之间能形成合适的氢键；从酸碱滴定的实验结果看，DNA 分子中氨基和酮基可能是形

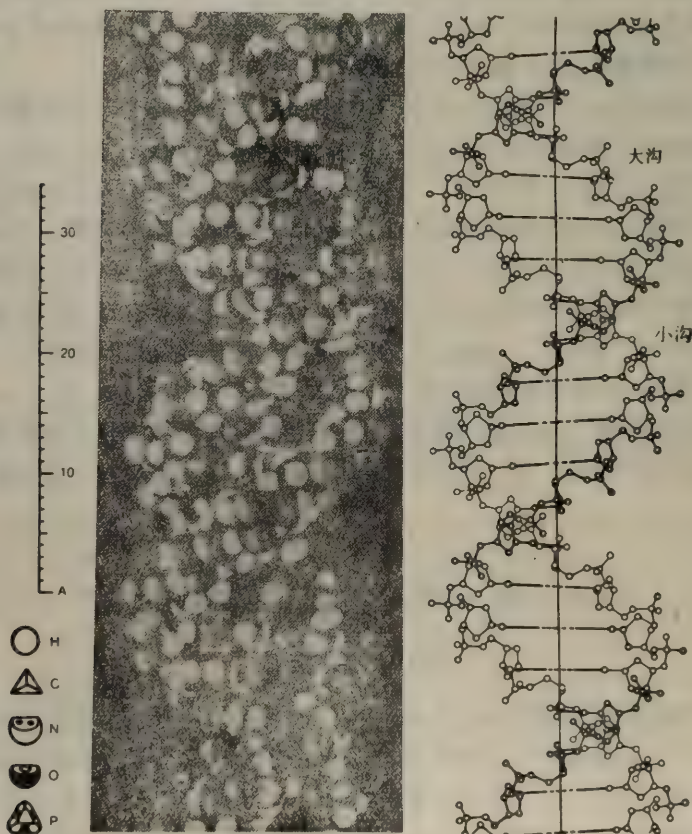


图 2-28 双螺旋 DNA 的分子模型

左为实心模型，右为相应的糖-磷酸主链，并标明大沟和小沟。

成氢键的。

如果我们把 DNA 想象成是双螺旋分子，那么以上的结果都可以得到完满的解释。按照这个假定去建立 DNA 分子模型，立即发现 DNA 必然是由二条链组成的右手双螺旋(图 2-28)。

双螺旋的基本特征

主链　脱氧核糖和磷酸基相互间隔连接构成 DNA 的主链。从化学键的方向来看，双螺旋中两条多核苷酸链是反平行的。二条主链处于螺旋的外侧，碱基处于螺旋的内部，由于糖和磷酸根的化学性质，主链是亲水性的。两条链形成右手螺旋，有共同的螺旋轴，螺旋的直径是 20 Å。

碱基对　由于几何形状的限制，只能由嘧啶和嘌呤配对才能使碱基对合适地安置在双螺旋内。若二个嘧啶配对则几何形状太小，二个嘌呤配对则几何形状又太大，为双螺旋所容纳不下。只有 A-T 碱基对和 G-C 碱基对的几何形状正适合双螺旋的大小。这两种碱基对有一个重要的特征，就是它们具有二次旋转对称性，即一对碱基对旋转 180°，并不影响双螺旋的对称性，因此双螺旋结构只限定了配对的方式，并不限定碱基的顺序(图 2-29)。

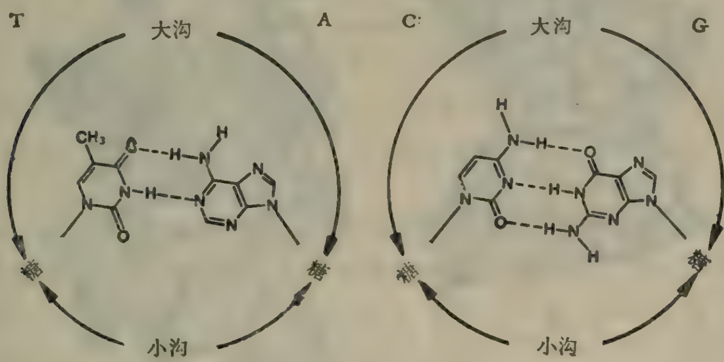


图 2-29　DNA 双螺旋中的碱基对

碱基环是一个共轭环，本身构成一个平面分子。在双螺旋中这个平面垂直于螺旋轴，相邻的两个碱基上下间隔 3.4 Å，每十对碱基组成一节螺旋，因此双螺旋的螺距是 34 Å。一条链中每个相邻的碱基的方向相差 36°。碱基之间的疏水作用可导致碱基堆积，这个引力同碱基对之间氢键一起稳定了双螺旋结构。

大沟和小沟　沿螺旋轴方向观察，配对的碱基并不充满双螺旋的空间。由于碱基对的方向性，使得碱基对占据的空间是不对称的，因此在双螺旋的表面形成二个凹下去的槽，一个槽大些，一个槽小些，分别称为大沟和小沟。双螺旋表面的沟对 DNA 和蛋白质的相互识别是很重要的，因为只有在沟内才能觉察到碱基的顺序，而在双螺旋的表面，是脱氧核糖和磷酸的重复结构，没有信息可言。

双螺旋的种类

上面我们所讨论的双螺旋的特征是属于所谓B型双螺旋。这是 DNA 钠盐的纤维在相对湿度 90% 时所处的状态。当外界条件变化时，双螺旋的特征也有变化，出现了新的形式。一般认为B型双螺旋最接近于 DNA 分子的溶液状态。

当 DNA 钠盐（或钾盐、铯盐）在相对湿度 75%时，DNA 纤维就处于A型的构象。它同B型的区别在于碱基平面不是垂直于螺旋轴，而是倾斜 20°，每个碱基升高 2.56 Å，二个相邻的碱基夹角为 32.7°，螺距是 28 Å，糖磷酸主链的构型也稍区别于B型（图 2-30）。

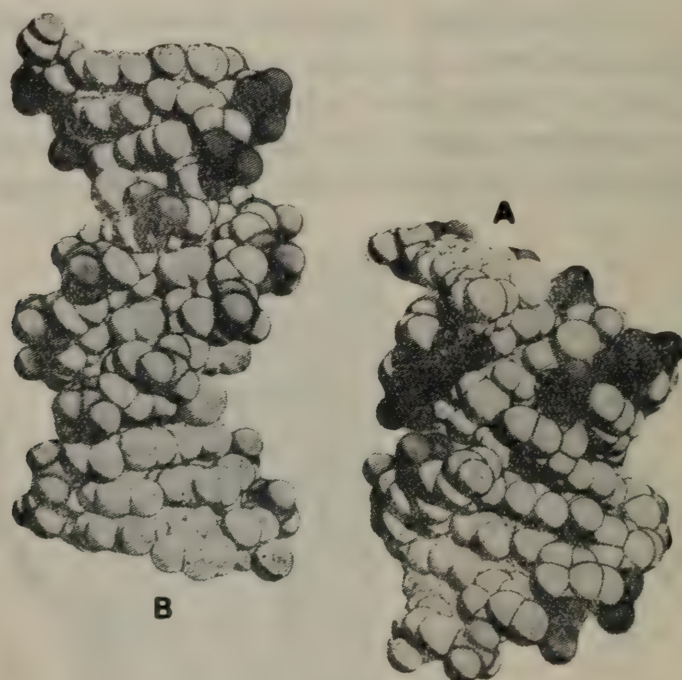


图 2-30 A-DNA 和 B-DNA 的比较

A型和B型是 DNA 分子的二个基本的双螺旋形式。在一些均聚或共聚核苷酸中，还可能出现一些其他的构象形式，但它们可以分别归入A型或B型。如 B′ 型可以在多聚 d(A)-多聚 d(T) 和多聚 d(I)-多聚 d(C) 中看到，它们的结晶形式稍区别于B型 DNA，但是其螺旋常数（$h = 3.25$ Å，$t = 36°$）还是属于B型的变种。当 DNA 是锂盐相对湿度降到 66%时，DNA 变成C型。原来认为C型和B型差别很大，研究表明并非如此。C型 DNA 中，碱基的倾角是 6°，每个碱基升高 3.31 Å，二个相邻碱基对的夹角是 38°，螺距是 31 Å，这些数据都和B型接近。近似于B型双螺旋的还有D型，这种型式在多聚 d(AT)-多聚 d(AT)、多聚 d(GC)-多聚 d(GC)、多聚 d(ATT)-多聚 d(AAI) 和多聚 d(AIC)-多聚 d(ICT) 等多聚物中出现。它同B型 DNA 在结构上的主要差别在于每个碱基仅升高 3 Å，倾角增加到 45°，小沟的宽度明显减小。在这种结构中，大分子

在外部不能从小沟中感觉到碱基顺序的变化。

属于A型一类的双螺旋还有 A′ 型和 A″. A′ 型只是在多聚 d(T)-多聚 d(A)-多聚 d(T) 三股螺旋中看到。 A″ 型只是在多聚 d(C)-多聚 d(I)-多聚 d(C) 中出现。三种A型中它们的 $C_{3'}$ 都是内糖环构型，就是 $C_{3'}$ 和 $C_{5'}$ 是在同一方向，因此在A型 DNA 中碱基对在螺旋轴之前。表 2-3 列出不同右旋双螺旋 DNA 的结构参数。

表 2-3 不同右旋双螺旋 DNA 的结构参数

双 螺 旋	碱基倾角	碱基夹角 *t*	碱基升高 *h*	螺距	每螺内碱基数	小沟宽 *m*	大沟宽 *M*	小沟深 *d*	大沟深 *D*
B-DNA	0	36.0°	3.37Å	34Å	10	5.7Å	11.7Å	7.5Å	8.5Å
B′-DNA		36.0°	3.28			6.9	13.5	6.5	8.5
C-DNA	6°	38.0°	3.31	31	9.3	4.8	10.5	7.9	7.5
D-DNA		45.0°	3.03			1.3	8.9	6.7	5.8
A″-DNA		30.0°	3.26			10.9	10.0	4.2	12.6
A′-DNA		32.7°	3.16			10.2	7.9	4.4	12.2
A-DNA	20°	32.7°	2.56	28	11	11.0	2.7	2.8	13.5

总的看来，A型螺旋比B型螺旋拧得紧一些，碱基的倾角大一些，大沟的深度大大超过小沟的深度。在B型螺旋中，大沟的宽度比小沟大，而两者的深度相差不远。

外界因素明显影响到 DNA 的构型。前面已经提到，相对湿度对 DNA 纤维的构型很有影响，在 90% 相对湿度时 DNA 呈B型，当相对湿度降到 75%时就变成A型。盐的种类和浓度对构型同样有明显的影响，当 NaCl 浓度较低时(<3%)，即使相对湿度仍在 92%，DNA 还是以A型存在；当 DNA 是钠盐时，A、B、C 三种构型都可能出现，而改成锂盐时，只有C型和B型可能出现。

另一个影响螺旋构型的因素是 DNA 的碱基组成和顺序以及顺序的重复性，其中起主要作用的是碱基的顺序。大多数 A-T 丰富的 DNA 自然倾向于形成B型双螺旋。但碱基顺序不同会发生构型改变，如多聚 d[AGT]-多聚 d[ACT] 虽然也是 A-T 丰富的，而它的构型类似于多聚 d[G]-多聚 d[C]；多聚 d[GC]-多聚 d[GC] 的构型不象它的组成类似物多聚 d[G]-多聚 d[C]，而象它的顺序类似物多聚 d[AT]-多聚 d[AT]。总的来讲，顺序和构型之间的关系仍是很不清楚的。

很明显构型的变化对 DNA 的生物功能将会有明显的影响，例如由于 RNA 分子存在 2′-OH，使得凡是含 RNA 的双链分子只能以A型的构型存在。当 RNA 聚合酶在 DNA 模板上合成 RNA 时，DNA 很可能必须变成A型。另外也有人讨论 DNA 的变构效应和长距离效应。例如在阿拉伯糖操纵子中，阻遏蛋白作用于离转录起始点 150bp 以外的位点上，即使距离如此远仍然能阻遏 RNA 聚合酶的转录作用。据猜测可能阻遏蛋白结合到 DNA 分子上去以后改变了 DNA 的构型，比如说碱基的倾角，这种效应可以传到启动子上，使得转录不能进行。又例如小牛胸腺 DNA 结合了一种抗菌素后，即使结合量只占 DNA 分子的 15%，但整个分子同溴乙锭结合的能力都变化了。可能 15% DNA 结构的变化引起了整个分子的构型的变化。总之，以上的迹象都提示了 DNA 双螺旋的构型不是僵化不变的。

双螺旋在溶液中

普遍认为 DNA 分子在水溶液中也是以双螺旋的形式存在，DNA 溶液的物理化学性质如变性和复性、粘度、圆二色性和小角度X光散射等结果支持这个观点。而且普遍认为在水溶液中 DNA 呈现B型构型，对这个假定也有一些实验支持。

近来有一些实验表明在水溶液中 DNA 的构型可能稍偏离于B型。从能量角度出发作理论计算，预言在溶液中 DNA 最稳定的构型的螺距应该是10.5个bp，而不是B型所规定的10个bp。用 ϕX174 DNA 复制型作电镜观察，也认为螺距应该是10.5bp，每个碱基升高2.9Å，既区别于B型的3.4Å，也区别于A型的2.55Å。

在溶液中，DNA 双螺旋的构型随条件而变化。例如在水溶液中加入酒精，双螺旋就会从B型变到C型，最后变成A型。B型是极性溶液中最稳定的结构。

双螺旋结构的研究动态

以上介绍了经典的 DNA 双螺旋模型的结构，概括地说这个模型有三个基本特征：DNA 的两条链是反向平行排列的；两条链的互补碱基对之间以氢键相联；两条链单向地按右旋方式互相绕在一起。双螺旋结构的这些性质已为大量实验事实所证明，可以认为是客观地反映了 DNA 的内在本质，因而是正确的。近年来，发现了 DNA 左手双螺旋构型，这是对 DNA 双螺旋结构的一个补充和发展。另外，有些科学家对 DNA 两条链互绕成螺旋的真实性提出了不同的见解。

A. Rich 等人用X射线衍射法分析人工合成的 DNA 小片段的晶体时，出乎意料地发现 d(CGCGCG) 是左旋的双螺旋。在这种双螺旋中两条六核苷酸分子链以反方向的形式排列着，其间互补 GC 碱基对以氢键互相联结，相邻碱基对互相平行。这些性质基本上与右旋 DNA 一样，引人注意的双链的旋转方向与右旋相反，而且磷原子联结线呈锯齿形，这种 DNA 被命名为 Z-DNA（图2-31）。左旋 Z-DNA 和右旋 B-DNA 的主要差异列于表2-4中。

Z-DNA 并不只限于 GC 相间的 DNA 顺序中，很可能嘌呤和嘧啶相间排列的顺序有助于 Z-DNA 结构的稳定性。通过从鼠和兔血中诱发的 Z-DNA 抗体，已经发现天然 DNA 中有的部分可以与 Z-DNA 的抗体相结合，因此 Z-DNA 的确是 DNA 的一种存在形式。

由于人工合成脱氧寡核苷酸的技术已日趋成熟，合成5—10个任意指定顺序的小片段 DNA 已不成问题(见第十章)，这就为进一步的结构分析提供足够的样品。因此近年来从人工合成的小片段 DNA 晶体结构分析获得了大量有关 DNA 结构的信息，从而证实了 A-DNA (CCGG) B-DNA (CGCGAATTCGCG) 和 Z-DNA (CGCG) 的存在。有趣的是在两段 CGCG 中插入 AATT 后的十二核苷酸片段以 B-DNA 的形式存在（见第十章表10-22）。这意味着一级结构对 DNA 双链的绕向及其空间结构有一定影响。

最近在测定 pBR322 DNA 的拓扑环绕数时发现了另一个用经典的双螺旋结构难以解释的现象。在4326对碱基组成的这种环状 DNA 中，若按每10对碱基组成一个右螺旋计算，则闭环状 pBR322 DNA 的拓扑环绕数应为 4×10^2，但实际测得的数值较此小得多，这说明在 pBR322 DNA 中可能同时存在着左旋和右旋的结构。

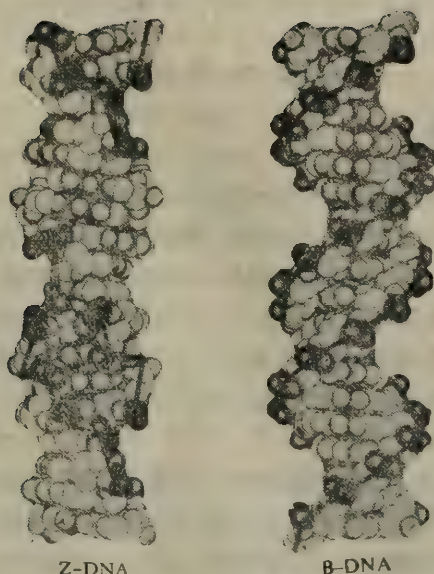


图 2-31 Z-DNA（左旋）和 B-DNA（右旋）的分子模型

表 2-4 B-DNA 与 Z-DNA 的比较

	B-DNA	Z-DNA
螺旋方向	右　旋	左　旋
每一转的核苷酸数目	10	12
直　径	~20 Å	~18 Å
每对核苷酸之间的轴向距离	3.4 Å	3.7 Å
螺　距	34 Å	45 Å
碱基平面的倾斜角	6°	7°
相邻碱基对之间的转角	36°	60°
糖苷键构象 d Gr	反　式	顺　式
d Cr	反　式	反　式
磷到轴心的距离 d GpC	9.0 Å	8.0 Å
dCpG	9.0 Å	6.9 Å
轴心与碱基的关系	穿过碱基对	不穿过碱基对

前几年新西兰的几位科学家提出可能存在一种“肩并肩”的 DNA 结构。在此结构中，DNA 分子的两条链不象经典双螺旋那样顺右旋方向绕在一起，而是绕了半周又返绕回去，如图 2-32 右图所表示。左图为经典双螺旋 DNA 模型，两条分子主链顺右旋方向绕在一起，其间以氢键维系；右图为“肩并肩” DNA 模型，两条分子的主链间同样以氢键维系，但并不绕在一起。他们认为这样可以比较容易解释 DNA 的复制机制。

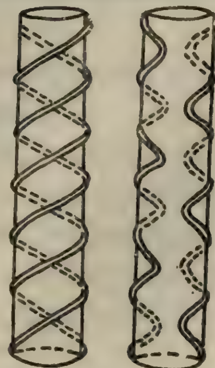


图 2-32 两种 DNA 结构示意图
左为经典的双螺旋模型，右为“肩并肩”模型。

1953 年 Watson 和 Crick 提出的 DNA 双螺旋模型已经 30 年了，但 DNA 的精细结构究竟如何？双螺旋的绕向与其一级结构关系怎样？双螺旋结构怎样受外界因素影

响？DNA 结构在复制与转录中的作用如何？等等，都需要进行更为深入的研究。

变性和复性

维持双螺旋分子的力

维持双螺旋分子的力有二种，一种是氢键，一种是疏水键。

氢键　氢原子同一些带电负性的原子形成共价键时，电子云偏向带电负性原子一边，氢原子带部分正电荷。这种氢原子同另一个带电负性的原子相互吸引而形成的氢键。氢键的受体可以是带部分负电荷的原子，也可以是带单位负电荷的原子。如果受体和供体都是带单位电荷的原子，那就是离子键了。离子键是氢键的一个特例。

在氢键中，氢原子都是以共价键连在氧原子或氮原子上。带电负性的受体大多也是氮原子和氧原子。氢键的键能在 3—7 千卡/克分子之间。受体和供体之间的电荷差别愈大，它们之间的键能也越大。氢键的强度比共价键弱得多，但比范德瓦尔斯力大得多。氢键不同于范德瓦尔斯力还在于前者有方向性，后者没有方向性，当氢原子直接指向受体原子时，其引力最大。若方向偏差一点，键能就小多了。氢键的受体和供体原子都有一定的特点，因此氢键的专一性比范德瓦尔斯力要强得多。

有关 DNA 双螺旋中的碱基之间的氢键前面已经讨论过了。还要指出的是 G-C 之间可以形成三个氢键，A-T 之间可以形成二个氢键。因此 G-C 配对的强度比 A-T 配对要强些。

疏水键　疏水键本质上是一种范德瓦尔斯力。当二个原子彼此接近时便产生诱导的振动电荷，由于这种电荷的分布而产生了非特异性的吸引力，这就是范德瓦尔斯力。很明显，范德瓦尔斯力没有分子的特异性，极性和非极性的分子都可以产生。这种力同作用基团之间的距离的六次方成反比，因此随着基团之间的距离减小，范德瓦尔斯力很快增加；另一方面，当距离太近时，由于外层电子云重叠，又引起了强大的斥力。斥力和引力达到平衡的距离称为范德瓦尔斯力半径，一般在 1.2—2.0 Å 之间（表 2-5）。范德瓦尔斯键能平均在 1 千卡/克分子左右，仅仅比分子的平均热能（在室温约 0.6 千卡/克分子）稍大一些。这意味着必须使分子的几何状态精确地配合，并且几个基团协同作用才能使范德瓦尔斯力有效地成为分子间的引力。

表 2-5　范德瓦尔斯半径

原　　子	半径(Å)
H	1.2
N	1.5
O	1.4
P	1.9
S	1.85
甲基	2.0
芳香环半厚	1.7

范德瓦尔斯力比氢键的能量弱得多，因此凡是能形成氢键的分子主要是形成氢键而不是形成范德瓦尔斯键。大家知道，水分子之间能形成较强的氢键，构成一种局部有序的

类似晶格的结构。因此假如一种不能同水形成氢键的化合物（比如苯）同水混合，水分子和苯分子虽能形成范德瓦尔斯力，但没有水分子本身的氢键能量大，因此水分子仍然自身维持在一起，而排斥苯分子的插入，所以水和苯不能互溶。一些极性分子，比如葡萄糖，它含有很多能形成氢键的基团，因此当葡萄糖插入到水的结构中去的时候，虽然破坏了水分子之间的氢键，但水和糖之间可以形成新的氢键。假如葡萄糖上一个氢原子被甲基取代，甲基比氢原子更严重地破坏水分子的结构，也就是说需要更多的能量才能把这个分子插入到水分子中间去。反之，与氢原子相比，水分子以更多的能量把甲基推出水的结构，这个能量约在 2—3 千卡/克分子左右。水分子排斥非极性基团的这种倾向经常称为疏水键。这个名称可能不太恰当，因为在这种情况下并没有产生新类型的键，水分子仍以氢键结合在一起，非极性键仍以范德瓦尔斯力结合在一起。但是这个词比较形象地表达了非极性分子有不愿同水接触而自己聚集在一起的倾向。很明显，一切破坏水的氢键结构的试剂也都能破坏疏水键。

DNA 分子的碱基都是由芳香环构成的，有很强的疏水性质。DNA 双螺旋在水溶液中，由于碱基有规则的排列，能形成较强的疏水键。有人估计疏水键对 DNA 双螺旋稳定性的贡献可能大于氢键的贡献。

破坏双螺旋的条件

上面讨论了维持双螺旋的力是由氢键和疏水键提供的，因此凡是破坏氢键和疏水键的因素都能导致双螺旋的破坏，如加热、极端的 pH、有机溶剂、尿素、酰胺等试剂。DNA 双螺旋的破坏称为变性。

DNA 的变性不仅受到外部条件的影响，而且也取决于 DNA 分子本身的稳定性，比如 DNA 分子中 G-C 含量高，分子就比较稳定。因为 G-C 之间有三对氢键，而 A-T 之间只有二对氢键。又比如环状双键 DNA 比相应的线状 DNA 稳定，因为拓扑学的原因，使得双螺旋不容易解开。

变性现象的宏观表现

双螺旋 DNA 分子具有一定的刚性，而变性之后就成为无规线团，伴随着变性过程 DNA 溶液的流体力学性质也发生变化。天然 DNA 的发色基团藏在双螺旋的里面，变性后碱基都暴露出来了，因此变性过程也伴随着 DNA 光学性质的变化。

由于以上原因，DNA 变性后表现为溶液的粘度大大降低，沉降速度增加，浮力密度上升，紫外吸收值升高。我们可以利用这些性质跟踪 DNA 的变性过程。

利用 DNA 变性后的增色效应跟踪变性过程是最方便的手段。如果用升高温度使 DNA 变性，以温度对紫外吸收值作图，得到一条 S 形曲线。在一个相当狭的范围内，增色效应是一个跳跃(图 2-33)，这同结晶的融化现象类似。产生紫外吸收值跃变的温度称为 DNA 的融点，通常是以增色效应达到一半的温度作为融点。在一定的条件下，融点的高低由 DNA 分子中的 G-C 含量所决定，融点和 G-C 的含量可用下列经验公式表示:

$$\%(G+C)=(T_m-69.3)\times 2.44$$

T_m 是 DNA 的融点。图 2-34 示出 T_m 与 G-C 含量的关系。

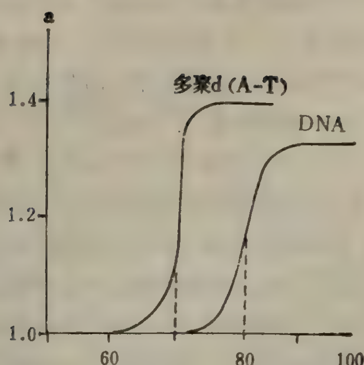


图 2-33 DNA 的变性曲线

与不含 G-C 的多聚 d(A-T)比较。纵坐标为相对的 260nm 吸收值，横坐标为温度。

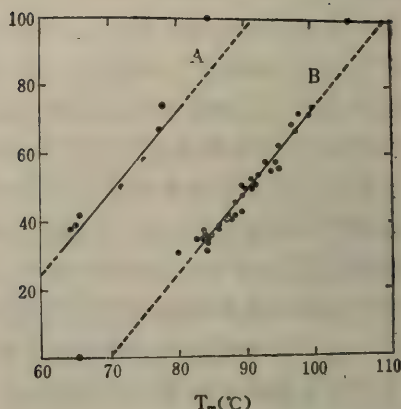


图 2-34 DNA 融点(T_m)与 G-C 含量的关系

A, 0.01M PO_4^{3-} + 0.001M EDTA;
B, 0.15M NaCl + 0.015M 柠檬酸钠。

融点不仅受 DNA 本身性质的影响，而且受溶液条件的影响。例如低离子浓度溶液中融点较低，融点范围也较宽；离子强度升高，转变范围也变狭。DNA 组成比较均一，它就在一个狭的温度范围内变性，如果 DNA 组成不均一，或者掺杂其他 DNA，融程就拉长了。

变性的微观机制

从 DNA 的变性曲线看，融程比较狭，反映了变性过程是一个协同的过程，就是在达到融点时 DNA 双螺旋是同时解开的。实际上由于 DNA 分子的碱基分布不均一，DNA 双螺旋的融化是一步一步进行的，是一个区域接着一个区域发生的，而不是协同地瞬时地进行。A-T 丰富的区域先融化，G-C 丰富的区域后融化，因此在精细的变性图上变性曲线不是一条平滑的 S 形曲线，而是包含了很多肩胛（shoulder）。如果把这条曲线对温度微分，就得到一系列小的峰，每一个峰代表一个小的融化区域，峰的高度大体等于融化区的碱基长度。因此可以从这些数据计算出各融区的大小、G、C 含量以及各个小融化区的排列次序。

复 性 动 力 学

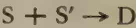
解除变性条件以后，变性的二条互补链可以重新结合起来，恢复成原来的双螺旋，这个过程称为复性。

很多因素影响复性的速度。简单顺序的 DNA 分子，如多聚 d(G)-多聚 d(C) 彼此发现互补顺序并不困难，因此很快就复性了。假如是一个真核 DNA，显然一些顺序在与互补链相碰撞之前，必然会碰到很多非互补链，因此在同样条件下，顺序复杂的 DNA 比顺序简单的 DNA 所需的复性时间要长些。同一种 DNA，当浓度高时，互补顺序相碰撞的机会增加，复性速度也快。DNA 片段的大小也影响复性的速度，因为大的线状单链扩散速度受到妨碍，减少了发现互补顺序的机会，因此在复性实验中，往往把 DNA 用机械剪力切成一定大小的片段。温度也影响复性的速度，低温不仅减少了互补链的碰撞机

会，而且使已经误配的片段再解开寻找正确的互补链变得困难。因此复性温度不能过低，一般比融点低 25℃ 左右是最佳温度。溶液的离子强度也影响到复性的速度，二个带很多同性电荷的 DNA 会相互排斥，因此在复性溶液中必须有一定浓度的盐。一般来讲，盐浓度高复性速度快。

复性过程是一个复杂的多步骤的反应过程。复性的第一步是当二条链相互碰撞时，假如在某个区域遇到正确的配对，就会先形成一个双链核心。核心形成后互补链就能很快地找到同伴，相互配对，完成复性过程。核心的形成是反应速度的限制因素。

复性过程可以用二级反应动力学公式来处理：



$$S + S' \rightarrow D$$

S 和 S′ 是 DNA 互补单链，D是双链 DNA。按二级反应动力学公式：

$$\frac{dc}{dt} = -kc^2$$

c 是 t 时间的单链 DNA 的浓度。积分后

$$\frac{C}{C_0} = \frac{1}{1 + C_0t \cdot K'}$$

C_0 是单链 DNA 的起始浓度，以每升溶液中核苷酸的摩尔数表示。K 和 K' 是速率常数。

因为在一个指定的实验中 C_0 是已知的，C 的浓度是可以测定的，如以 $\frac{C}{C_0}$ 对 C_0t 作图，得到如图 2-35 所示的曲线，在曲线的中点，即

$$\frac{C}{C_0} = \frac{1}{2} \text{时，} \quad C_0t_{1/2} = \frac{1}{K'}$$

如前所述，很多因素会影响复性的速度，但是我们如果把所有的可变因素，包括温度、离子强度、片段的长短都控制起来，那么剩下的唯一可变因素就是顺序的复杂性。实验指出，速度常数 K' 与 DNA 中的复杂性 N 成反比：

$$K' = \frac{1}{N}$$

例如，大肠杆菌 DNA 比 λ 噬菌体 DNA 分子大 100 倍。这样，大肠杆菌 DNA 复性速度也比 λDNA 要小 100 倍。因此 $C_0t_{1/2}$ 的大小代表了基因组的大小，每个基因组的顺

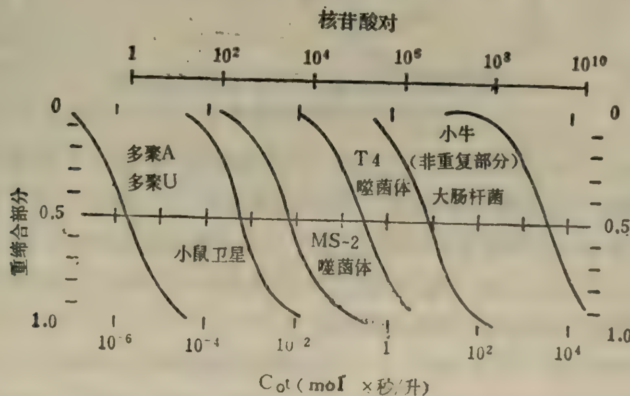


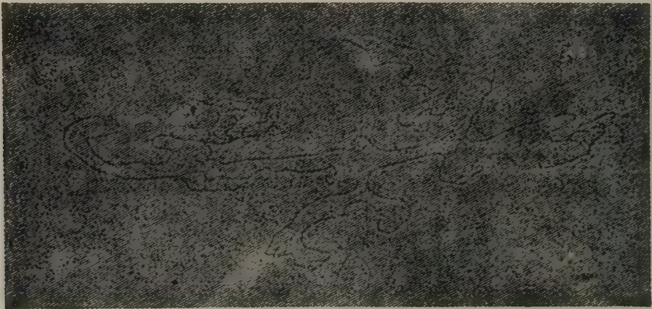
图 2-35 DNA 的复性曲线

序都不一样，因此 $C_0t_{1/2}$ 也代表了 DNA 顺序的复杂程度。

如果用这种方法测定真核 DNA 的 C_0t，按理它的 C_0t 数值是很大的，但是实验结果发现 $\frac{C}{C_0}$ 对 C_0t 作图并不是一条单纯的 S 形曲线，而可以看作几个 S 型曲线的加和。比如鼠的 DNA 中 10% DNA 的复性速度甚至比已知最小的病毒 DNA 复性还要快，因此说明这部分是由很多短的重复顺序所组成。对于真核生物 DNA 顺序的组织的知识就是从复性动力学开始获得的。

应　　用

在 DNA 研究工作中，很多手段都是同 DNA 的变性和复性有关的，前面已经提到了一些这方面的工作。



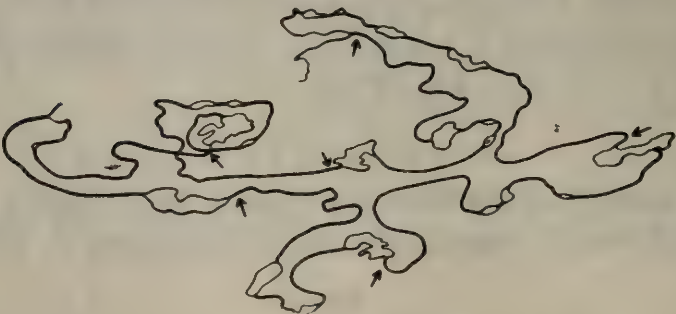


图 2-36　环状爪蟾核糖体 DNA 的部分变性图
箭头表示重复区的边界

前面已经提到，DNA 变性是一个区域一个区域发生的，A-T 丰富的区域最容易变性，电子显微镜下确实也看到这种情况。在一定的条件下，DNA 分子长链上容易变性的部分呈现一个个小泡，而没有变性的部分保持一条双链(图 2-36)。每一种 DNA 的局部变性图谱都是不一样的，因此可以利用这个手段识别 DNA 分子中某种基因的位置，或者作为一个参照位置用来确定其他功能的位点。

杂交方法广泛应用于 DNA 研究中。杂交的基础就是 DNA 的变性和复性，因此应该注意影响变性和复性的各种因素。利用杂交方法可以求出特定基因的频率、异源 DNA 的相似程度，配合电镜方法可以了解基因的组织、基因的定位。近几年发现真核 DNA

中有插入顺序最早就是借助于 DNA 和 RNA 的杂交方法。

四、DNA 的三级结构

DNA 的三级结构指双螺旋链的扭曲。超螺旋是 DNA 三级结构的一种形式，DNA 在核小体结构中的扭曲方式也是一种超螺旋结构。 核小体可以组成更高层次的结构，DNA 分子也将进一步扭曲，因此 DNA 三级结构还可以分成若干层次。

超　螺　旋

在 DNA 双螺旋中，每 10 个核苷酸长度旋转一圈，这时双螺旋处于能量最低的状态。如果这种正常的双螺旋 DNA 额外地多转几圈或少转几圈，就会使双螺旋内的原子偏离正常的位置，在这种双螺旋中存在额外的张力。如果双螺旋的末端是开放的，这种张力可以通过链的转动而释放出来，DNA 将恢复到正常的双螺旋状态。如果 DNA 分子的二端是以某种方式固定的，或者是环状分子，这些额外的张力就不能释放到分子之外，而只能在 DNA 内部使原子的位置重排，这样 DNA 分子本身就会扭曲，这种扭曲就称为超螺旋。

为了说明超螺旋 DNA 的性质，人们常用拓扑学的研究成果来表示两个互相绕在一起的环在几何形态上的相互关系。当两个连续不断的环状分子互相绕在一起而又不能用物理方法分开时，这种分子的性质就可以用拓扑环绕数(LK)予以描述（见图 2-37)。当两个环中的一个放在某一平面上，另一个环就要从这个平面的上下穿过，图 2-37 的 (a)、(b) 和 (c) 分别表示另一个环穿过平面环的一、二和十次，这样，它们的拓扑环绕数分别为 1、2 和 10。图 2-37 中的(d)和 (c) 的意义是完全相同的，只是(d)表示两个环分别围着一个公共轴互相绕在一起。如果令此公共轴保持在一个平面上，则绕在一起的两个环的投影就一定有偶数交点，此偶数应为拓扑环绕数的一倍。只要两个环状分子都没有发生临时的或永久的断裂，双环之间的拓扑环绕数就是一个恒定不变的整数，永远不会发生变化。因而这个原理适用于由两条互补单链组成的环状 DNA 分子。

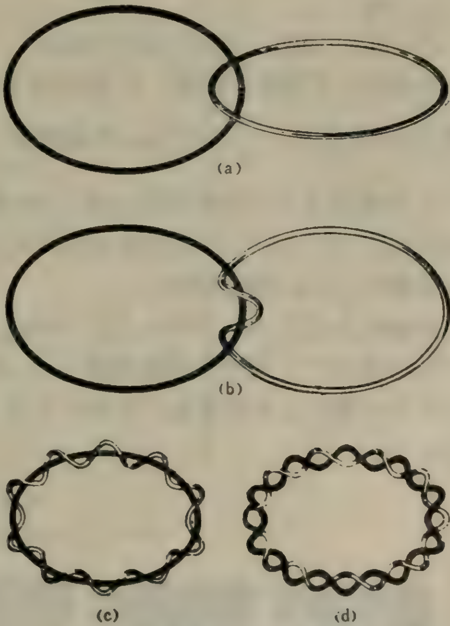


图 2-37　拓扑环绕数示意图(c 和 d 的结构在拓扑学上是相同的)

拓扑环绕数尽管是两个闭合环状线段之间相互关系的重要数据，但是，有相同拓扑环绕数的闭环物却可以有种种不同的空间关系。图 2-38 表示 $Lk = 0$ 时出现的情况，(a)表示拓扑环绕数为 0 的两个环平行地排在一起，(b)表示引入三个右螺旋时，必须要同时引入三个左螺旋才能保持 $Lk = 0$，(c)和 (d) 分别表示在 (b) 中引入的三个左螺旋可以

表现为空间盘绕的形式，也可以表现为双链互绕的形式。很明显，在(c)和(d)中双链的公共轴都绕着处于同一平面上的超螺旋轴转了三次。

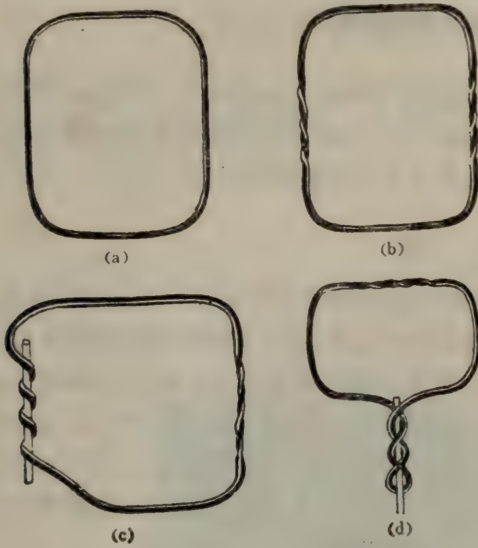


图 2-38 拓扑环绕数为 0 时闭环双链的空间关系

为了说明两个闭合环状线段之间的空间关系，数学家已经证明：$\alpha=\beta+\tau$。其中 α 即为拓扑环绕数 Lk，β 为双链环绕数，又称全绕率 (total twist)。在核酸化学中，$\beta=+1$ 定义为双链向右旋 360°。τ 为超螺旋数或拧数(writhing number)。这个公式与常见于文献的 $Lk=T_w+W_r$ 意义相同。

由上述公式可直接导出 $\tau=\alpha-\beta$，其含意为：拓扑环绕数与双链环绕数之差即为超螺旋数。当 $\tau=0$ 时，环状双链 DNA 就处于松弛态；$\alpha>\beta$ 时，环状分子呈正超螺旋；当 $\alpha<\beta$ 时，则为负超螺旋。实际上所有的已知环状双链 DNA 的超螺旋数值都是负的，即其拓扑环绕数总是小于双链环绕数。以含 5226 核苷酸对的 SV40 DNA 为例，按B型 DNA 双螺旋模型，其 $\beta=\frac{5226}{10}=522$，实际测得 $\tau=-26$，因此 $\alpha=\tau+\beta=-26+522=496$。说明正是由于双链之间少绕了 26 圈才使整个 DNA 分子内产生一定张力，驱使双链闭环的 SV40 DNA 以超螺旋的形式存在。当其中有一个缺口或 260 对碱基间的氢键遭受破坏，此张力就会消失。

超螺旋数目是以一个 DNA 分子为单位的。DNA 分子有大有小，比如带 10 个超螺旋对大肠杆菌 DNA 的性质可能影响不大，而对 SV40 DNA 的影响可能很明显，因为前者分子量比后者大三个数量级。因此为了便于比较引入超螺旋密度这个概念，超螺旋密度的定义为：

$$\sigma=\frac{\tau}{\beta^0}$$

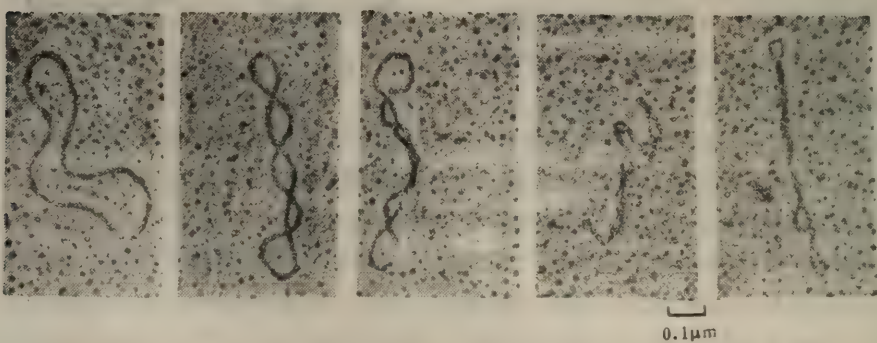


图 2-39 小 Col EI 质粒二倍体 DNA (5kb)的不同超螺旋形态的电子显微镜图

σ 为超螺旋密度，β^0 是 DNA 分子的碱基对数目的十分之一。

线状 DNA 或环状 DNA 都可以有三级结构，但对于研究工作利用小分子环状 DNA 是最方便的，因为环状分子在体外容易保留超螺旋状态。尽管线状 DNA 在体内也有超螺旋，但在体外我们没有办法把它的两端固定，因此在分离过程中内部的张力都释放掉了。以下提到的有关超螺旋 DNA 的性质、测定等内容都是用小分子环状 DNA 作为实验材料获得的。图 2-39 是环状 DNA 的电子显微镜图。

由于共价闭环双链 DNA 没有可以自由转动的末端，因此需要有更高的温度，也需要较高的 pH 才能破坏超螺旋 DNA 中的碱基对。例如超螺旋的多瘤病毒 DNA 融点达 140°，而它的复制型 II DNA 的融点只有 89°。超螺旋分子的结构比较紧密，因此它的粘度比较低，浮力密度比较大，沉降速率比同样大小的线状分子快 20%。

有些试剂如溴乙锭、二氨基吖啶、放线菌素 D 等可以插入 DNA 双螺旋碱基对之间的空间中去，从而改变超螺旋的状态。DNA 结合了溴乙锭以后，降低了 β 数，使负的超螺旋减少。随着溴乙锭量的增加，负超螺旋不仅可以完全松开，甚至可以进一步引入正的超螺旋，测定超螺旋的方法之一就是利用溴乙锭引起超螺旋变化这一特性。

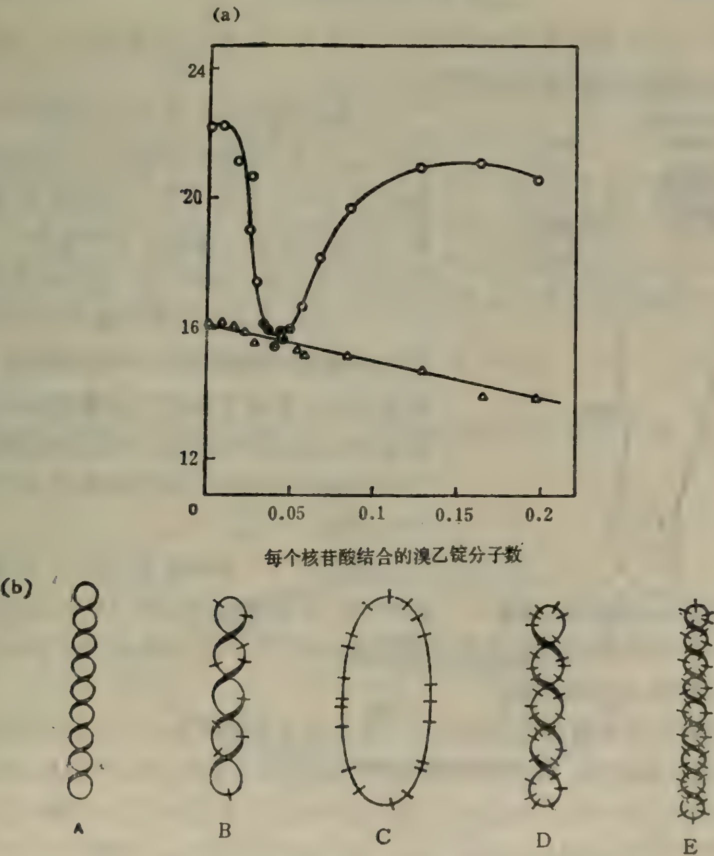


图 2-40 溴乙锭对多瘤病毒 DNA 沉降常数的影响及其解释

(a) ○表示超螺旋态，△为非超螺旋态，●只形成单一沉降区带的 DNA 分子。(b) 为(a)实验的解释。双螺旋以单一的连续线表示。从 A 至 E 说明由于溴乙锭的结合，使得负超螺旋减少(B)，完全松弛(C)，并引进正超螺旋(D 和 E)。

上面已经提到,同松弛形 DNA 相比较，超螺旋 DNA 的沉降常数大一些。如果在不同浓度的溴乙锭溶液中测定一个超螺旋 DNA 的沉降常数(S),将会发现,随着溴乙锭的增加 S 值先是下降,达到一个最低值以后 S 值又上升(图 2-40)。S 值最小的状态就是 DNA 完全松弛的状态。从处于松弛态 DNA 结合溴乙锭的数量可以推算分子内包含多少超螺旋,因为每结合一个溴乙锭能使双螺旋解开 26° 左右(以前认为每结合一个溴乙锭能使双螺旋解开 12°，现在看来这个数值大大偏低了）用粘度方法可以代替超离心的方法,原理是一样的,技术上可能更方便一些。我们注意到，加入溴乙锭以后 S 值出现一个低谷,这意味着该 DNA 分子中包含了负的超螺旋，如果 DNA 分子中包含的是正的超螺旋,将不会出现这样一个低谷。现在发现所有天然 DNA 中的超螺旋都是负的。

由于超螺旋的 DNA 分子比较紧密，在凝胶电泳中的电泳速度比较快。利用这个性质也可以测定超螺旋的数目。用拓扑异构酶 I 温和作用于超螺旋的 DNA，可以产生一系列带有不同超螺旋数目的中间体。这些中间体可以用电泳方法分开，每个中间体只相差一个超螺旋数目,因此只要计数电泳图上中间体的数目就可以知道超螺旋的含量。

生物体内的超螺旋

生物体内的 DNA 都具有某种形式的三级结构。比较小的双链环状 DNA，如病毒和质粒 DNA 在体内都以负超螺旋的形式存在。

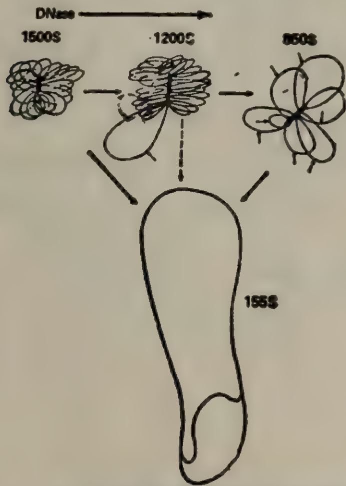


图 2-41　大肠杆菌染色体的结构示意图

大肠杆菌 DNA 是环状双链分子，整个分子长达 1mm 多，在体内压缩成 1μm 直径的小体。大肠杆菌 DNA 包含了几个独立的超螺旋。据推断，折迭的大肠杆菌染色体可能有一个 RNA 核心，核心外面是 DNA 环，DNA 环的数目在 12—80 个之间，每个环都构成一个独立的超螺旋。用脱氧核糖核酸酶轻微作用，可以仅松弛一个超螺旋环而不影响其他环的超螺旋结构，但是再用核糖核酸酶处理可以把所有的超螺旋打开，按照这些实验结果推测大肠杆菌染色体的结构如图 2-41 所示。

核小体　真核生物 DNA 主要存在于染色体中。染色体除了包含 DNA 之外，还包含了组蛋白。对 DNA 染色体复杂的构型以前了解得很少。自从发现了核小体之后，DNA 在染色体内的构型逐步明朗起来。

染色质丝在电子显微镜中呈念珠状。一根长丝上附着很多小球,小球的直径为 100 Å 左右,小球之间是一根细丝，小球和细丝互相间隔重复无数次(图 2-42)。每一个重复单位称为一个核小体。每个核小体包含了 200 对碱基和五种组蛋白。在细丝上结合有组蛋白 H1，其他四种组蛋白(H2A，H2B，H3 和 H4)位于小球内。如果把小球两边的细丝用核酸酶消化掉,留下来的小球称为核小体的核心。核心的 DNA 长约 140 碱基对，这一段 DNA 绕在由组蛋白 H2A，H2B，H3，H4 各二个组成的八聚体蛋白核心之外。从不同生物获得的核小体 DNA 长短有所区别，从 140 到 240 碱基对，而核小体核心中

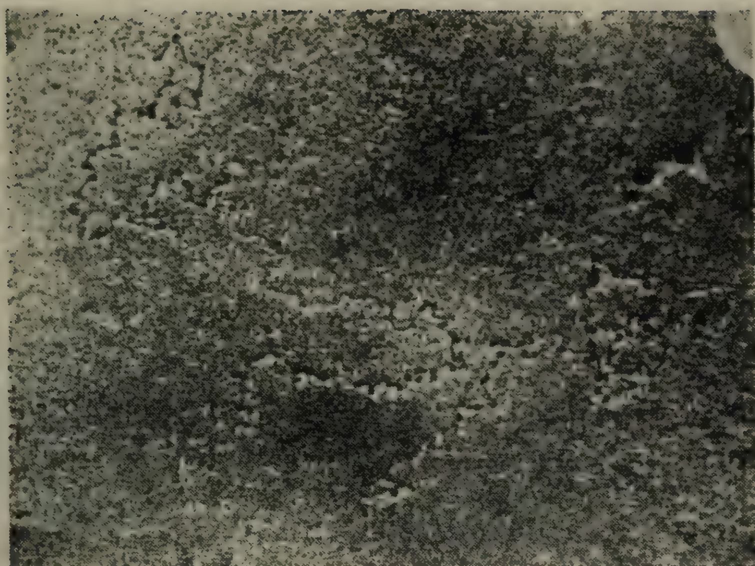


图 2-42 染色质丝的电子显微镜图

DNA 的长度几乎总是 140 碱基对。

从电子显微镜看，核小体的核心好象是一个圆球体。用中子散射和X射线衍射方法测定溶液中核小体的外形，发现核小体不是球形而是短粗的圆柱体，直径 100 Å，高 50 Å。更直接的证据来自对核小体核心结晶的研究，用X射线衍射配合电子显微镜，测定核小体核心的分辨率可达到 20—25 Å 水平。这些结果表明，核心是平盘状的，直径 110 Å，高 57 Å，这和中子散射的结果是一致的。在这个分辨率水平，还不能跟纵 DNA 链的走向，但是从中子散射的实验可以肯定，DNA 双螺旋是绕在组蛋白八聚体之外，因此可以推测 DNA 双螺旋形成 90 Å 直径的超螺旋，绕在蛋白核心之外，一共绕 $1\frac{3}{4}$ 圈，螺距为 28 Å，这和B型双螺旋可形成每圈 75—82 碱基对的超螺旋的推算是一致的（见图 2-43）。

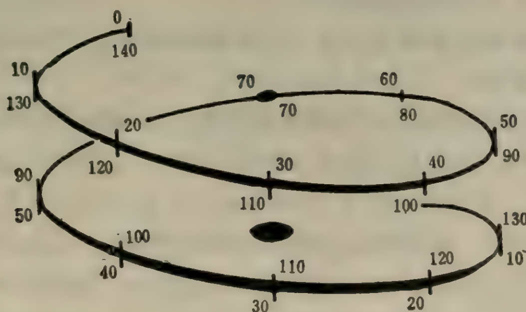


图 2-43 核小体结构示意图

当中黑点表示由 H2A、H2B、H3 和 H_4 各二分子组成的八聚体蛋白核心，外面为由 140bp 围绕的 DNA 分子，构成一个核小体核心。

DNA 在染色体内的高次结构　在核小体内 DNA 被压缩了 7 倍。而在染色体内

DNA 双螺旋压缩了将近 10,000 倍，那么核小体如何进一步压缩而构成染色体呢？这方面还没定论。但是有些迹象表明染色体是多层次的螺旋结构。

染色体细丝的直径约 100 Å，现在可以肯定，染色体细丝是由核小体串连而构成的。染色体的粗丝直径约 300 Å，中子散射的结果提示，粗丝是一种螺线管的形式，外径 300 Å，螺距 100—110 Å，中间有 100 Å 的空洞。因此可以推测粗丝是由细丝绕成的螺旋，粗丝每一螺旋包含了 6 个核小体。当溶液中存在镁离子时，用电子显微镜可以看到这种 300 Å的结构，如果没有镁离子，只能看到 100 Å 的细丝，因此镁离子对螺线管的形成是重要的。另外组蛋白的 H1 可能对核小体的压缩也有关。

从核小体到螺线管，DNA 双螺旋压缩了 6 倍，进一步的压缩可能是通过形成所谓超螺线管而达到的。有迹象表明，中期染色体是一个长的中空的圆筒，直径约 4,000 Å。这个圆筒是由直径 300 Å 的螺线管绕成的，压缩比是 40。这个超螺线管称为单位纤维。单位纤维本身还可以压缩 5 倍。这样总的压缩比是 $7 \times 6 \times 40 \times 5$，同总的 DNA 压缩比接近(图 2-44a)。

也有人认为核小体细丝先形成发夹形状的环，这种环进一步绕成一个螺旋，形成染色体结构(图 2-44b)。

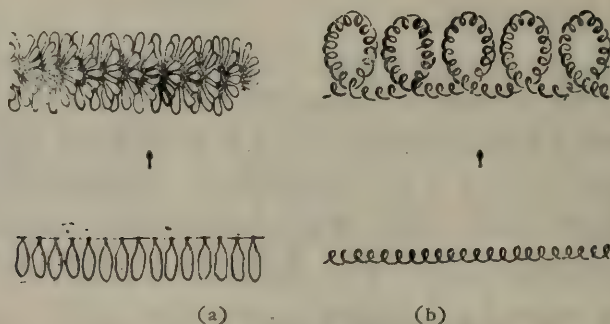


图 2-44 DNA 在染色体中的高层次结构

有一点可以肯定，染色体看来是一层层螺旋绕起来的。DNA 双螺旋就在这些有层次的结构中绕缠，形成有层次的超螺旋。

DNA 超螺旋有结构上的作用，更可能有功能的意义。但是这些功能的意义还有待开发，这里只提几个迹象。

很多病毒的复制都要经过具有超螺旋的双链环状 DNA 阶段。例如噬菌体 ϕX174 是一个单链环状 DNA，在复制过程中 ϕX174 DNA 先复制成双链环状 DNA，然后由病毒自身编码的 A 蛋白在双链 DNA 的特定位点上切一刀，使得 ϕX174 DNA 能以滚筒复制的方式复制出来。有趣的是 A 蛋白作用的底物必须是具有超螺旋的闭环双链的 DNA，如果用同样闭环双链 DNA，而不具有超螺旋，A 蛋白就不起作用。

如果人为地把一段回文结构的 DNA 片段引入一个双链环状分子中，再向这个环状分子引入负的超螺旋。这样在电镜中看到在这个环状分子上长出二个“触角”，这是在回文区形成的发夹结构(图 2-45)。因此超螺旋可能有助于正常细胞中回文结构从规则的螺旋中突出来，成为一个容易识别的信号。图中所示的三个人工制造的闭环 DNA 中都有很长的回文结构，当用拓扑异构酶 II 引入较多的超螺旋后，整个分子由于超螺旋引起

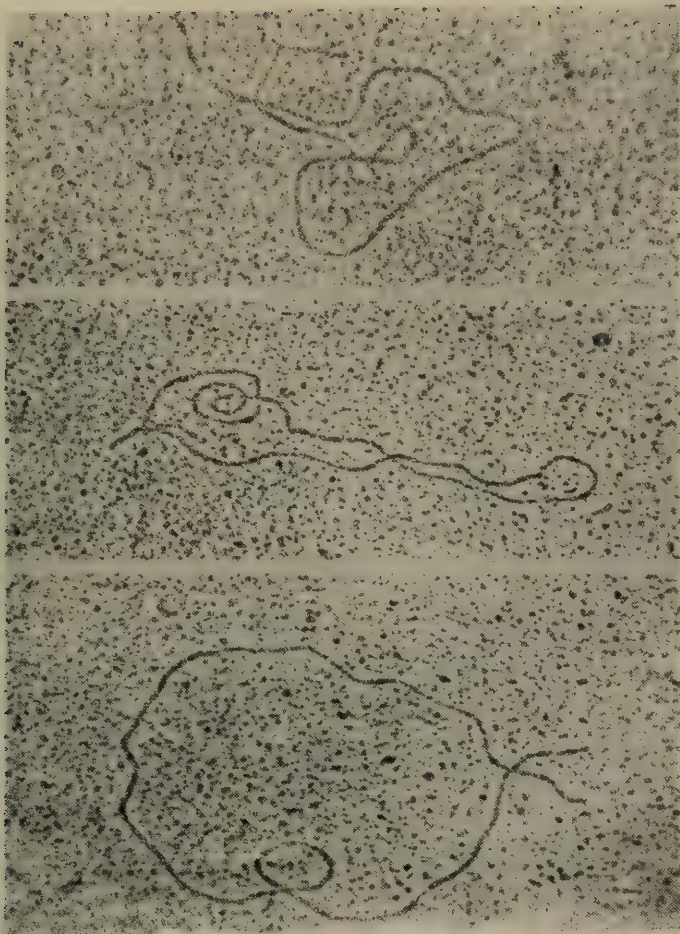


图 2-45 带回文结构的超螺旋 DNA

的张力会消失在回文结构形成的发夹结构中。

超螺旋可能对转录的控制也有关系。SV40 DNA 如果具有超螺旋，那么它在体外转录是不对称的；而松弛型的 SV40 DNA 在体外转录是对称的。

在 ϕX174 复制型 DNA 上，转录时转录的 RNA 同双链 DNA 形成一个 R 环，这个环的形成需要超螺旋释放的自由能。这样超螺旋就成了一种调控因素，因为在一个启动基因上消耗了超螺旋的自由能之后，就妨碍了其他启动基因的转录，也妨碍了 DNA 的复制。

参 考 文 献

[1] Adams, R. L. P. et al. (eds.), 1981, Biochemistry of the Nucleic Acids, 9th ed. London, Chapman and Hall.

[2] Freifelder, D. (ed.), 1978, The DNA Molecule Structure, and Properties (original papers, analyses, and problems) W. H. Freeman and Company, San Francisco.

[3] Maxam, A. M. and Gilbert, W., 1980, Sequencing End-Labeled DNA with Base-Specific Chemical Cleavages, in *Methods in Enzymology*, 65. 499—560, eds. Grosman, L. and Moldave, K. Academic Press.

[4] Sanger, F. and Coulson, A. R., 1975, A Rapid Method for Determining Sequences in DNA by

Primed Synthesis with Polymerase, *J. Mol. Biol.*, **94**, 442—448.

[5] Sanger, F. et al., 1977, DNA Sequencing with Chain-terminating Inhibitors, *Proc. Natl. Acad Sci.* (USA) **74**, 5463—5467.

[6] 洪国藩 (Hong, G. F.), 1982, A Systematic DNA Sequencing Strategy, *J. Mol. Biol.* **158**, 539—549.

[7] Nathans, D. and Smith, H. O., 1975, Restriction Endonucleases in the Analysis and Restructuring of DNA Molecules, *Ann. Rev. Biochem.* **44**, 273—293.

[8] Watson, J. D. and Crick, F. H. C., 1953, Molecular Structure of Nucleic Acids. A Structure of DNA, *Nature* **171**, 737—738.

[9] Wang, A. H.-J. et al., 1979, Molecular Structute of a Left-Handed Double-Helical DNA Fragment at Atomic Resolution, *Nature*, **282**, 680—686.

[10] Kornberg, R. D. and Klug, A., 1981, The Nucleosome, *Scientific American*, **244**(2), 48—60.

第三章 RNA 结 构

吴仁龙 祁国荣

核酸是生物体的重要高分子化合物，它储存着生物体全部遗传信息，是基因表达不可缺少的基础物质。除 DNA 外，在生物体中还存在着另一类核酸——RNA。RNA 有许多种(详见表 3-1 和 3-2)，它们具有不同的生物功能。

表 3-1 大肠杆菌中 RNA 的含量和性质

种 类	沉降系数	分 子 量	核苷酸残基数	占细胞总 RNA 的百分比	代谢性质
rRNA	5S	3.5×10^4	120	82%	稳定
	16S	5.5×10^5	1542		稳定
	23S	1.1×10^6	2904		稳定
tRNA	4S	$2.5—3\times10^4$	74—93	16%	稳定
mRNA	4-25S	$2.5\times10^4—10^6$	70—3,000	~2%	不稳定
部分其他 RNA	4.5S		107		稳定
	6S		184		稳定
	10S		609		稳定
	4.5S		109		不稳定
	4.5S 前体		130		不稳定

成熟 RNA 细胞内含量最多的几类 RNA 分子的主要生物功能是参与蛋白质的生物合成。第一类是核糖体 RNA (rRNA)，原核生物中有三种 (5S rRNA、16S rRNA 和 23S rRNA)，真核生物中有四种 (一般是 5S rRNA、5.8S rRNA、18SrRNA 和 28S rRNA)。这些 RNA 分子在代谢上十分稳定，是构成生物体内蛋白质合成“机器”——核糖体的重要成分。第二类是转移 RNA (tRNA)，生物体内存在着与 20 种氨基酸相对应的 tRNA 分子，它们在代谢上也是稳定的，在蛋白质生物合成的过程中起着接受、转运和参入氨基酸的作用。参与蛋白质合成的第三类 RNA 分子叫做信使 RNA (mRNA)，它的生物功能是从 DNA 上把遗传密码即蛋白质中氨基酸排列顺序的信息接受过来，并起着模板的作用合成蛋白质。一般讲，一种蛋白质就有一种与之对应的 mRNA 分子，但有的 mRNA 分子可翻译出几种蛋白质，mRNA 在代谢上是不稳定的。rRNA、tRNA 和 mRNA 参与蛋白质生物合成的细节可参看蛋白质生物合成一章。这几类 RNA 都存在于细胞质部分，它们行使生物功能的主要场所在核糖体上。在真核细胞的线粒体和叶绿体里存在着独立的蛋白质合成系统，也有这几类 RNA 分子，但是其大小和结构往往与细胞质中的不完全相同。此外，生物体内还存在着多种具有特定生物功能的 RNA 分子，例如，参与启动 DNA 复制的引物 RNA；在某些肿瘤病毒中存在着一些类似 tRNA 的分子是反转录酶的引物，参与 DNA 的合成；在某些细菌中有的 tRNA 分子作为氨基酸

表 3-2 真核生物中 RNA 的存在和分类

位 置	名 称	沉降系数	分子量	核苷酸数	含 量*	代谢性质
细胞质	rRNA	5S	3.5×10^4	~120	50%	稳定
		5.8S	5×10^4	~160		稳定
		18S	6.5×10^5	~2000		稳定
		28S	1.7×10^6	~5000		稳定
	tRNA	4S	$2.5—3\times10^4$	~80	24%	稳定
	mRNA					不稳定
细胞核	tRNA 前体				11%	
	rRNA 前体	45S	4.1×10^6			
		41S	3.1×10^6			
		32S	3.1×10^6			
		20S	9×10^5			
	mRNA 前体		$1\times10^5—2\times10^7$			极不稳定
	snRNA	4—8S		80—260		稳定
线粒体或叶绿体	tRNA	4S(3S)			15%	
	rRNA	类似原核生物或更小				
	mRNA	类似原核生物				

* 为哺乳动物肝细胞中的含量。

的载体而参与细胞膜的合成，例如表皮葡萄球菌中的 $tRNA^{Gly}$ 在细胞壁合成时是甘氨酸的载体。还有一些 RNA 分子是一些酶活性不可缺少的组成成分，例如，兔肌 1，4-α-葡聚糖分支酶含有 31 个核苷酸的 2.5S RNA 分子，特异性地作用于 tRNA 前体的 RNase P 中有 77%的成分是核糖核酸(详见第五章)。真核细胞的核内还存在着多种代谢稳定的小分子 RNA (snRNA)，其大小从 4S 到 8S 不等，大约由 80 到 260 个核苷酸构成，其中有些 snRNA 在 RNA 成熟过程中起重要作用。另外，还有一些与染色质结合的 RNA (chRNA)，chRNA 的生物功能目前尚不十分清楚，有关这方面的细节可参看第八章。

前体 RNA　绝大多数 RNA 分子都是在细胞核内合成的。核内存在着各种 RNA 的前体，例如，tRNA 前体、rRNA 前体和 mRNA 前体，详见第五章。这些前体 RNA 的分子量一般要比对应的成熟 RNA 大得多，例如，哺乳动物肝细胞的 rRNA 的前体是 45S 细胞核 RNA (nRNA)，在成熟过程中产生 41S nRNA、32S nRNA 和 20S nRNA 等中间物，最后成为 28S、5.8S 和 18S rRNA。前体 RNA 都是由细胞核 DNA 转录产生的，经过剪切、装配和修饰等一系列加工过程以后产生成熟的 tRNA、rRNA 和 mRNA，进入细胞质中行使其生物功能。有人把在细胞核内合成后离开细胞核去行使生物功能的 RNA 分子称作迁移性 RNA (migrating RNA)，而始终存在于核内的叫作非迁移性 RNA (nonmigrating RNA)。有关细胞核内 RNA 的细节可参看第八章。

病毒 RNA　病毒中的遗传物质有的是 DNA，也有的是 RNA。绝大多数植物病毒和 RNA 肿瘤病毒以 RNA 作为遗传物质。一般讲，病毒 RNA (除负链 RNA 病毒外，包括噬菌体 RNA) 既是遗传信息的载体，也具有 mRNA 的功能。类病毒 (viroid)

是高等植物的一类病原体，它是一个没有外壳蛋白的游离 RNA 分子，分子量约 12 万。关于病毒 RNA 的细节也参看第八章。

这一章着重介绍细胞质中与蛋白质生物合成有关的 RNA 的结构和与测定 RNA 一级结构有关的一些问题。

一、RNA 的结构

tRNA

1965 年 Holley 等首先测定了酵母丙氨酸 tRNA 的核苷酸排列顺序以来，对各种不同来源的 tRNA 的一级结构进行了广泛和深入的研究。据不完全统计，迄今已测定一级结构的 tRNA 有二百五十多种。在原核生物中以大肠杆菌 tRNA 最多，除脯氨酸外其余 19 种都已测定。在真核生物中，酵母的二十种氨基酸对应的 tRNA 都已测定。根据目前已经测定的 tRNA 一级结构的资料，经过归纳和分析它们有以下共同的特征：组成 tRNA 的核苷酸通常为 74—93 个，其中大多数为 76 个；几乎所有 tRNA 都具有三叶草式的二级结构(见图 3-1)。

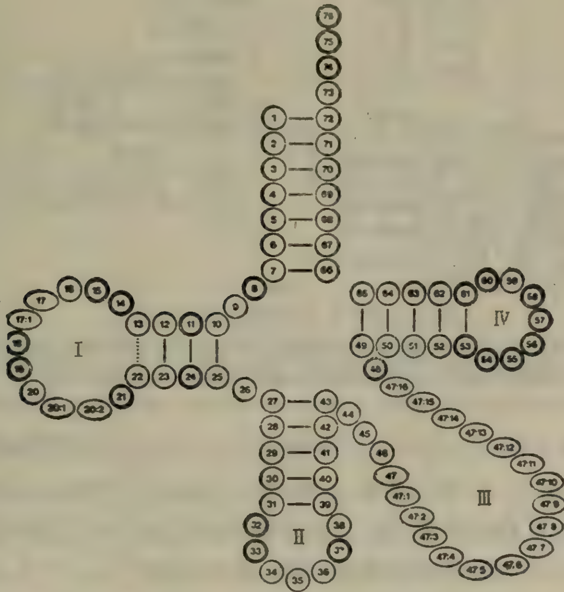


图 3-1 tRNA 二级结构通式(式中的核苷酸排列次序系按照酵母苯丙氨酸 tRNA 的次序)

◯表示所有 tRNA 在这个位置上都有核苷酸存在； ○表示所有 tRNA 在这个位置上都是固定不变的，或者变化较少；○表示并非所有 tRNA 在这个位置上都有核苷酸存在。在 tRNA 中加入核苷酸，其命名法如下：以加入的核苷酸前一个核苷酸的序号加冒号后再编号。例如 20：A 和 20：B 表示在第 20 位核苷酸之后加入的第一个和第二个核苷酸。I，二氢尿嘧啶环(D 环)；II，反密码子环；III，可变环；IV，TψC 环(简称 T 环)。

tRNA 的一级结构

不变和半不变的核苷酸——从噬菌体、细菌、真菌、昆虫、植物、哺乳动物(以及真核生物细胞器)和人的体细胞等的 tRNA 分子一级结构的资料分析，约有二十多个位置上的核苷酸是不变和半不变的。其中十多个位置上的核苷酸在几乎所有 tRNA 中都是不变的，

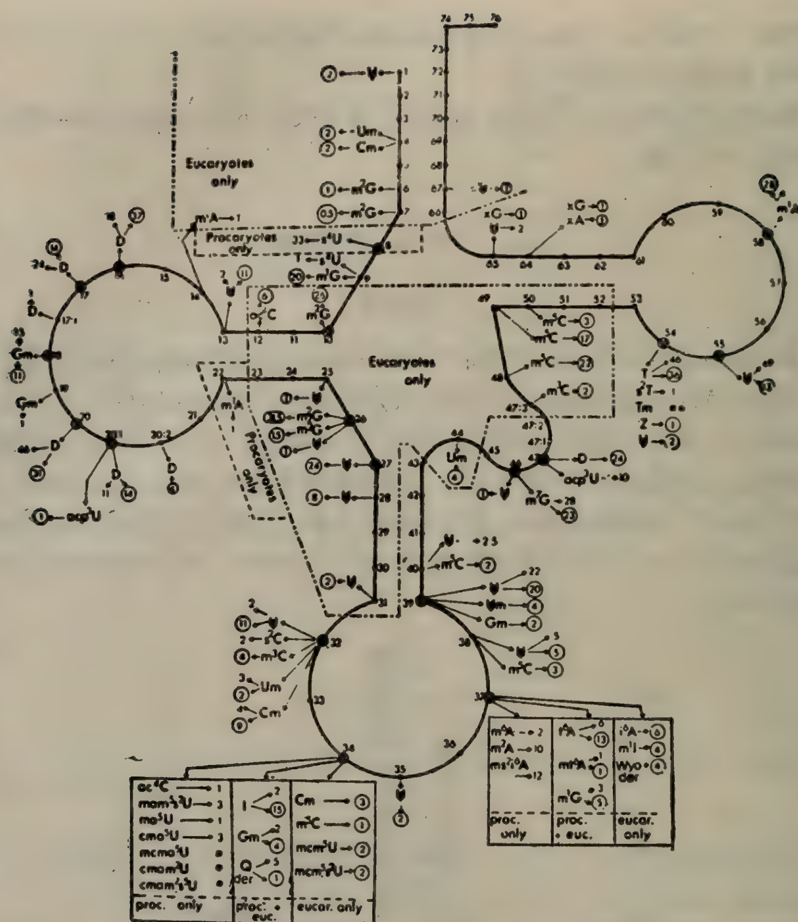


图 3-2 tRNA 中修饰核苷酸的分布

修饰成分符号见第一章。数字代表在该部位出现特定修饰核苷的已知 tRNA 数
(圆圈内的数为真核生物的,没有圆圈的数字为原核生物的
O Eucaryotes:真核生物, Eucar.: 真核, Procaryotes: 原核生物, Proc.:原核)

例如 3′ 端 CCA,第 8 位的 U(或U的衍生物),18 和 19 位的 G(或G的衍生物),53 位的 G,61 位的C和 58 位的 A(或A的衍生物)等。所谓半不变的核苷酸则是那些只可能发生嘌呤间或嘧啶间的互换的核苷酸。这些不变和半不变的核苷酸对于维持 tRNA 倒L型的高级结构和实现其生物功能起着重要作用,同时也说明了 tRNA 分子在进化上的保守性。

TψCG 序列——在绝大多数 tRNA 的第 54 到 57 位存在着 TψCG 序列。该区域称为 TψC 环。TψCG 对于 tRNA 分子与 5S rRNA 的结合和 tRNA 高级结构的维系起着重要作用。但是在少数 tRNA 中不存在这一序列。已知的大多数真核生物起始 tRNA 上这一序列被 AψCG 或 AUCG 所取代。一些具有特殊功能的 tRNA 分子,例如作为肿瘤病毒反转录酶的引物的 tRNA 分子,该处是 ψψCA, TmψCA 或 ψψCG; 在表皮葡萄球菌中参与细胞壁合成的 $tRNA^{Gly}$ 该区域为 UGCG 等。

修饰核苷酸——tRNA 分子中含有大量修饰核苷酸,也叫稀有核苷酸,迄今已发现近七十种稀有核苷酸中近五十种存在于 tRNA 中(见第一章)。每个 tRNA 分子中少则含

有2个,多则含有19个修饰核苷酸，它们大多分布在分子的非配对区。特别在反密码子3′端邻位出现修饰核苷酸的频率最高(见图3-2),而且大多属于修饰的嘌呤核苷酸,这对于反密码子环三级结构的稳定、反密码子第三字母和mRNA上密码子第一字母间严格而准确的结合是重要的。在反密码子第一位置或叫做摆动位置上,也常常是修饰核苷酸(见图3-2)。在密码子与反密码子相互作用时摆动位置的配对情况参阅第七章。迄今对于修饰核苷酸在 tRNA 分子存在的生物学功能了解还很少。

表3-3 tRNA 二级结构的证据

方　　法	结　　果	
1.紫外吸收光谱	酵母 tRNA 有50%双链	
2.红外吸收光谱	大肠杆菌 $tRNA_f^{Met}$ 配对碱基占53%	
3.旋光色散光谱	酵母 $tRNA^{Ala}$ 双链占50%	
4.核磁共振	$tRNA^{Phe}$ 有21±3碱基对	
5.互补核苷酸片段结合	能结合部位	不能结合部位
大肠杆菌 $tRNA^{Tyr}$	$ACCA_{OH}$ D环 反密码子环 可变环	所有茎区 T环
酵母 $tRNA_f^{Met}$	$ACCA_{OH}$ 反密码环	T环 D环
6.化学修饰的研究	可修饰	不可修饰
酵母 $tRNA^{Phe}$	CCA_{OH} D环 反密码子环 可变环	所有茎区 T环
7.自由能的计算及其分布的研究	对62种 tRNA 的一级结构研究结果三叶草型二级结构具有最低的自由能	

tRNA 的二级结构

tRNA 分子的三叶草式二级结构也是 Holley 等人在1965年首先提出的,当时的主要依据是酵母丙氨酸 tRNA 的一级结构信息和 Watson-Crick 碱基配对原则。根据这个模型,酵母丙氨酸 tRNA 分子中有20到21对碱基对，即占该 tRNA 中总数53%的核苷酸处于碱基配对之中。随后许多物理和化学(包括大量一级结构的信息)性质的研究证明，在溶液中的 tRNA 分子确实存在着与三叶草模型为数相当的碱基对。 这些方法包括紫外吸收光谱、红外吸收光谱、旋光色散特性、核磁共振、互补核苷酸片段结合、化学试剂修饰和在多种状态下自由能的计算等的研究,见表3-3。tRNA 二级结构模型也已经由酵母丙氨酸 tRNA 的X射线衍射分析的结果所证实。

在三叶草构型中(图3-1) tRNA 分子一般可分为氨基酸接受茎、二氢尿嘧啶茎(D茎)、反密码子茎、TψC 茎、某些 tRNA 还有可变茎、二氢尿嘧啶环(D环、I)、反密码子环 (II)、TψC 环(IV)和可变环(III)。其中接受茎通常有7对碱基对,反密码子茎5对碱基对，TψC 茎5对碱基对;反密码子环7个核苷酸和 TψC 环7个核苷酸在各种 tRNA

中保持不变。由此可见，各种不同的 tRNA 分子在长度上的变化主要发生在三个区域，即D环和可变环的核苷酸数目以及D茎的碱基对数目的不同。为了便于研究，人们根据 tRNA 分子可变区域的特征，把 tRNA 分成三种类型：第一类 tRNA，D茎 3 对碱基对，可变环 3—5 个核苷酸 (D_3V_{3-5})；第二类 tRNA，D茎 3 对碱基对，可变环 13—21 个核苷酸 (D_3V_{13-21})；第三类 tRNA，D茎 4 对碱基对，可变环 3—6 个核苷酸 (D_4V_{3-6})。从已经测定了一级结构的 tRNA 的结构类型来看，第三类结构 (D_4V_{3-6}) 的 tRNA 约占总数的 50% 左右。迄今为止发现的一个例外是哺乳动物线粒体丝氨酸 tRNA 只有 63 个核苷酸，呈现二叶草式二级结构(见图 3-3)，其沉降系数为 3S (一般 tRNA 为 4S)，这一结果也从 tRNA 基因的结构分析得到进一步确证。

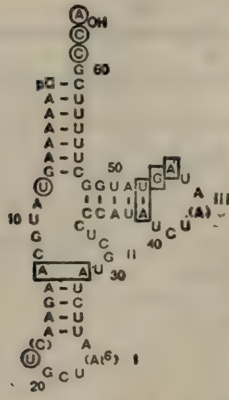


图 3-3 牛心脏线粒体丝氨酸 tRNA 二级结构 ○，不变核苷酸；()，半不变核苷酸；□，与一般 tRNA 不同的地方。

tRNA 的三级结构

70 年代以来 RNA 立体结构的研究有了很大的进展，先后有人利用许多物理学和化学的方法来测定 tRNA 等的立体结构。这些方法有激光拉曼光谱研究、碱基特异性化学修饰研究、游离嘌呤氚交换试验、荧光能量转换的研究等等。但是除了少数方法能够提供较多的结构信息包括核糖磷酸骨架的走向外，其余方法所提供的结构信息是局部的和零碎的，因而不能构画出 tRNA 分子清晰的三度空间结构。直到 1973 年—1975 年间，由于制备适合于高分辨率 X 射线衍射分析的 tRNA 结晶方法的突破，Kim、Rich 和 Robertus 等人利用这一方法，分别以两种不同的晶型(斜方晶和六方晶)测定了酵母苯丙氨酸 tRNA 的三度空间结构，得到了几乎完全一致的结果，提出了 $tRNA^{Phe}$ 分子的倒 L 型三级结构模型 (见图 3-4)。随后还有几种 tRNA (大肠杆菌起始 tRNA，酵母起始 tRNA 和大肠杆菌精氨酸 tRNA 等) 先后用同样方法测定了三级结构。tRNA 是三级结构了解较详细的一种天然存在的具有特定生物功能的 RNA 分子。

根据 3 Å 分辨率的 X 射线衍射分析结果，酵母 $tRNA^{Phe}$ 分子具有像字母 L 那样的折叠，有下面这些特征：(1)分子的全貌像字母 L。氨基酸接受茎和 TψC 茎组成一个双螺旋，D茎和反密码茎形成另一个近似连续的双螺旋，这二个双螺旋构成了像字母 L 的形状。L 的一端是 3′ 端 CCA，另一端是反密码子，两端之间的距离约 70 Å。TψC 环和D环构成 L 的拐角。整个分子是扁平的，厚度约 20 Å。L 型分子中的螺旋类似 DNA 分子中的 A 型结构。(2) 分子中存在着维系三级结构的氢键。大多数维系三级结构的氢键是在 tRNA 分子的许多不变和半不变的碱基与构成 tRNA 二级结构的碱基之间形成的(详见图 3-4C)。除有标准的 Watson-Crick 配对外，还有许多非标准的配对(见图 3-5)。图 3-5 中 a 和 b 是标准 Watson-Crick 配对；c、d、e_1、e_2、f_1、f_2、g、i、h_1 和 h_2 是非标准配对。(3) 维系三级结构的氢键中包含有碱基与核糖或磷酸之间形成的氢键。例如，A_{21} 的 N_1 与 U_8 的 $O_{2'}$；U_{33} 的 N_3 与 A_{36} 的 $O_{5'}$ 或磷酸上的氧形成的氢键；ψ_{55} 的 N_3 与 A_{58} 的磷酸上的氧之间形成的氢键等等。后二种情况的氢键分别稳定了反密码环和 TψC 环上尖的弯曲。在 DNA 中没有 $O_{2'}$，因而不能与之形成氢键，这也许是 DNA 分子往往是长链的一个原

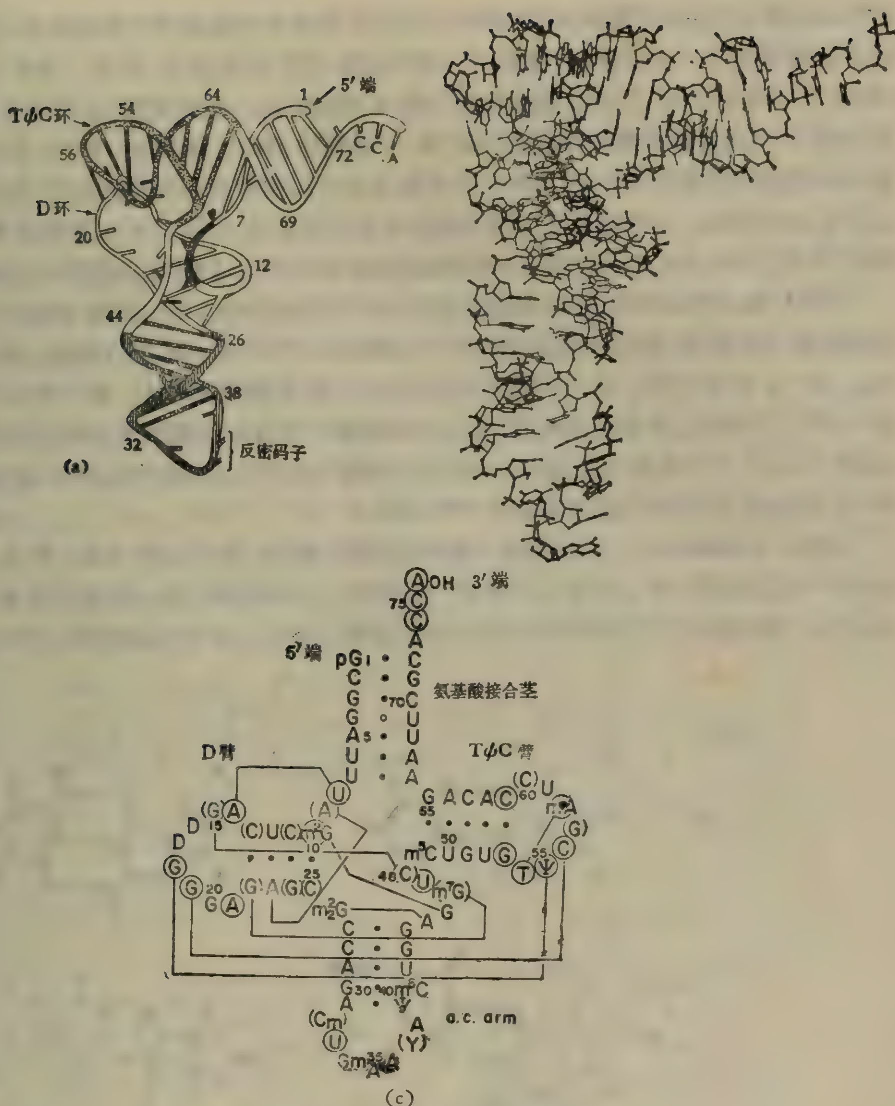


图 3-4　酵母苯丙氨酸 tRNA 的三级结构

(a) 链走向图；(b) X射线衍射分析得到的立体结构模型；(c) 分子中维持二级和三级结构氢键的位置。

因。(4) $tRNA^{Phe}$ 的三级结构具有假双重对称性。$tRNA^{Phe}$ 分子上所有的碱基几乎都有围绕着并包含在二个螺旋结构之中的趋势。约占分子 40% 的结构是对称的。在氨基酸接受茎和 TψC 茎与反密码子茎和反密码子环的一部分包括D茎之间有一个假双重对称轴，即在二个螺旋的交接处的中心有一个双重对称轴。

tRNA 三级结构的共同特征——酵母苯丙氨酸 tRNA 三级结构中发现了许多维持三级结构折叠的氢键大多数是由不变和半不变的碱基参与形成的。因此自然可以推测 L 型的折叠是所有与 $tRNA^{Phe}$ 具有类似二级结构特征的 tRNA (D_4V_5) 所共有的。目前已经测定了三级结构的几种 tRNA 包括起始 tRNA (大肠杆菌起始 tRNA 和酵母起始 tRNA) 和延伸 tRNA (大肠杆菌精氨酸 tRNA) 等都具有类似酵母苯丙氨酸 tRNA 那样的倒 L

型结构。例如，大肠杆菌精氨酸 tRNA (D_3V_5) 的X射线衍射分析结果显示，尽管组成它的核苷酸有四十二处与酵母 $tRNA^{Phe}$ 的不同，它的二级结构类型 (D_3V_5) 也与酵母 $tRNA^{Phe}$ (D_4V_5)不同，但其三级结构仍具有L型的折叠。由此可见，L型的折叠可能是所有 tRNA 包括真核 tRNA 和原核 tRNA 以及延伸 tRNA 和起始 tRNA 所共有的。当然仔细观察已经得到的几种 tRNA 分子的三级结构模型可以发现它们之间在精细结构上还是有差别的，这包括分子上两个螺旋形成的拐角的大小；CCA 末端伸展的程度；D环在分子上的松紧程度；反密码子臂的构象和反密码子环上的 U_{33} 碱基的方向等等。

tRNA 高级结构与其生物功能的关系——1975年 Kim 根据酵母 $tRNA^{Phe}$ 结晶的X射线衍射分析结果，提出了氨酰 tRNA 合成酶与 tRNA 的对称辨认假说，他认为合成酶辨认 tRNA 时分二步。第一步是一般辨认；第二步是特异性辨认。第一步是由合成酶辨认 tRNA 分子的L型骨架结构，即通过与假双重对称分子上相当大的部分的结合实现的(见图 3-6)。这个假说可以说明大肠杆菌的氨酰 tRNA 合成酶可以酰化某些真核生物 tRNA 的现象，但目前还缺乏直接的实验证据。

tRNA 与 5S rRNA 的相互作用。原核生物的 tRNA 分子上的 TψC 环上有一段序列 GTψC 与其同源的 5S rRNA 分子上的一段序列 GAAC 相互补。已有证据表明 tRNA 与 5S rRNA 相互作用发生在 tRNA 处于核糖体的A位上。许多证据包括 tRNA 高级结构

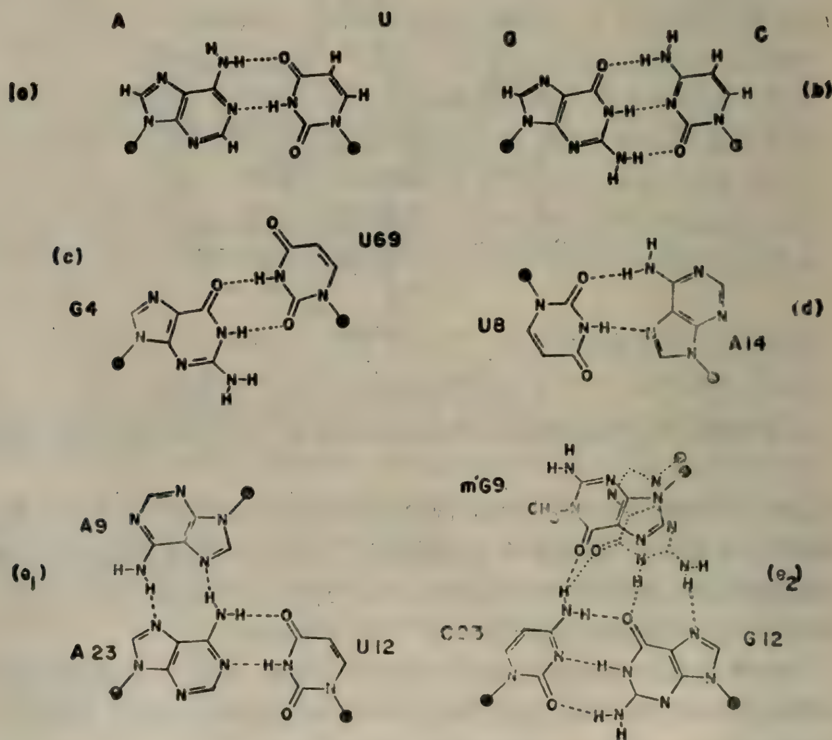
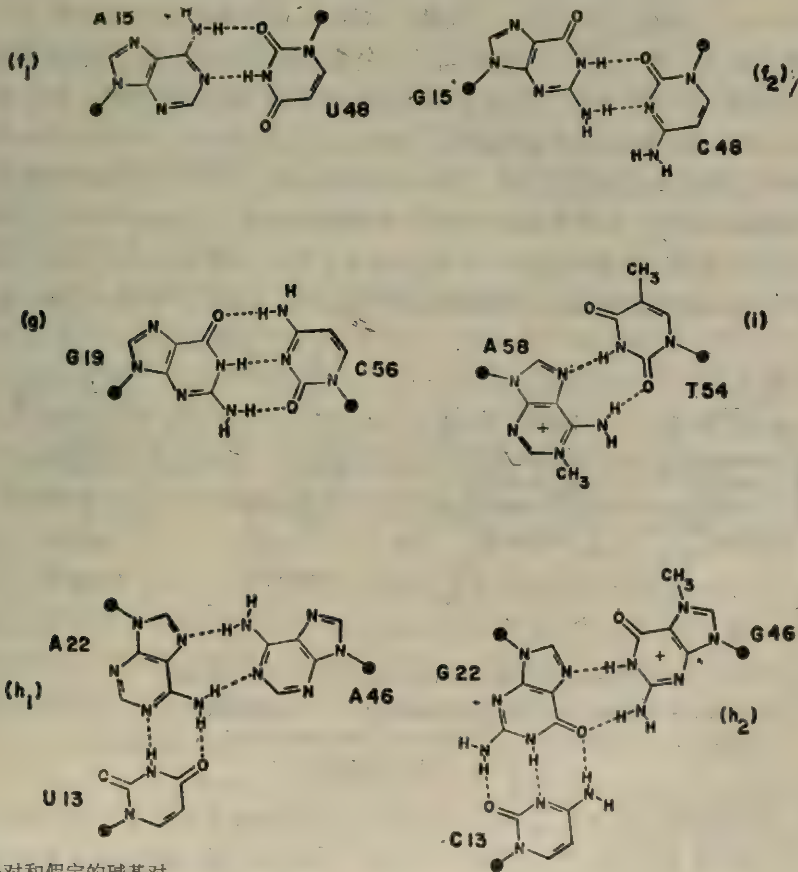


图 3-5 酵母苯丙氨酸 tRNA 中
a 和 b 为标准 Watson-Crick 配对；c、d、e1、f1、

测定的结果显示：GTϕC 是被 tRNA 分子内的氢键维系着的，或者具有严格的碱基堆积构象，即它像 tRNA 分子上茎区的碱基一样，化学上是不活泼的(见表 3-3)。因此只有通过核糖体或 mRNA 的诱导使 tRNA 发生构象的变化才能与 5S rRNA 相互作用。从酵母苯丙氨酸 tRNA 结晶的 X 射线衍射结果得知 G_{19} 与 C_{56} 之间的碱基对比正常的 GC 对的稳定性差，二个碱基不在同一平面上，形成 30° 的扭角。5S rRNA 与 tRNA 相互作用的可能的机制如图 3-7 所示。首先 5S rRNA 上的 CCGAAC 中的 G 或另外一个 G 残基与 C_{56} 形成另一个扭角碱基对，产生一个双碱基对的中间物(图 3-7, b)，这个碱基对降低了 $C_{56} \cdot G_{19}$ 碱基对的稳定性，并最后打开了它，转而触发了 G_{57} 的解堆积，至少拆开二个维持三级结构的碱基对 $\phi_{55} \cdot G_{18}$ 和 $T_{54} \cdot A_{58}$，可能还有 $G_{53} \cdot C_{61}$. 这样 T 环与 D 环解开(图 3-7, c)，GTϕC 与 5S rRNA 上的 GAAC 即可形成互补的碱基对。

在蛋白质生物合成中延伸 tRNA 与起始 tRNA 的重要区别或许在于氨酰化的 $tRNA_f^{Met}$ 是直接进入核糖体的 P 位，而氨酰化的延伸 tRNA 则首先进入核糖体的 A 位，然后再转移到 P 位。Rich 等根据大肠杆菌 $tRNA_f^{Met}$ 的 X 射线衍射分析结果，提出了延伸 tRNA 和起始 tRNA 功能上的不同所依赖的结构基础或许在于 U_{33} 碱基在反密码子环上的不同方向，即所谓的 U_{33}"转动假说"。Rich 等发现大肠杆菌 $tRNA_f^{Met}$ 上 U_{33} 的碱基投影在反密码子环的外面，而 $tRNA^{Phe}$ 的则在环内。尿嘧啶的不同方位可能允许 33 位的



的部分碱基对和假定的碱基对

t_2、c_2 g、i、h_1 和 h_2 是非标准配对。

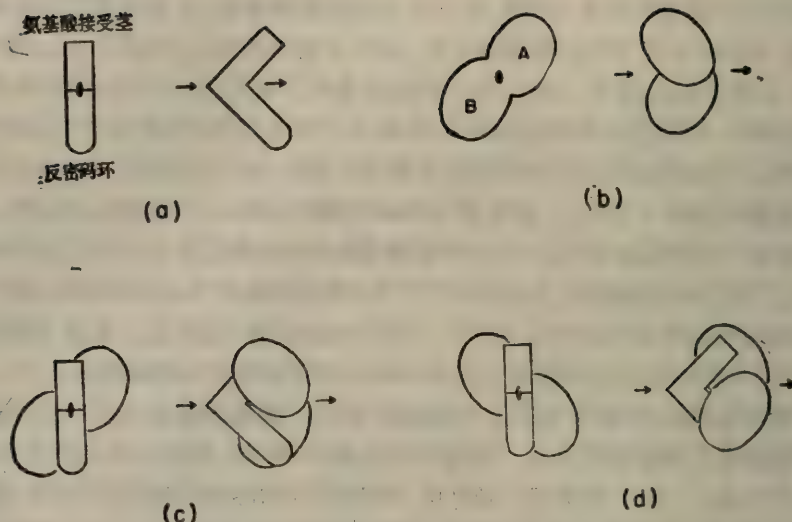


图 3-6　tRNA 与氨酰 tRNA 合成酶相互作用的模型

(a) tRNA 分子的构象；(b) 氨酰 tRNA 合成酶的构象；
(c)、(d) tRNA 与氨酰 tRNA 合成酶相互作用的可能的模式。

核糖的 $O_{2'}$ 与 36 位的磷酸之间形成一个氢键。这样 36 位的磷酸被拉向环的中心，这与正常的磷酸所受的向外的力不同，于是形成了一个歪斜的反密码子环，而这种构象与酵母 $tRNA^{Phe}$ 是不同的。这种构象的不同导致反密码子碱基排列的轻微差别，这可能就是允许起始 tRNA 首先进入核糖体 P 位的原因。Rich 认为，真核起始 tRNA 的 33 位虽是胞嘧啶，这不妨碍它获得与大肠杆菌起始 tRNA 相同的构象，其实核酸酶降解的研究已经表明大肠杆菌起始 tRNA 与真核起始 tRNA 有相同的构象。延伸 tRNA 在从核糖体上的 A 位向 P 位转移时会与核糖体上的一个元件相互作用，结果反密码子环从一种构象弹向另一种使 U_{33} 碱基向外的构象。如果是这样的话，U_{33} 可能在 tRNA 占据 P 位时起一

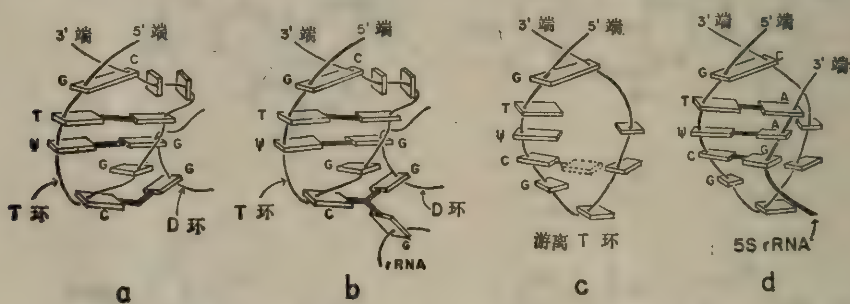


图 3-7　tRNA 与 5S rRNA 相互作用的一个假设的示意图

a，tRNA 分子上的 T 环与 D 环之间的碱基对；b，核糖体 rRNA 上的 G 残基与扭角 G.C 对相互作用；c，T 环与 D 环之间的碱基对打开；d，5S rRNA 与 tRNA 之间形成的碱基对。

种结构作用。在 mRNA 与 A 位及 P 位上的 tRNA 形成三维复合物时，U_{33} 上的 $O_{2'}$ 可能与在 A 位上的 tRNA 的 2′OH 形成氢键，同时两个反密码子与 mRNA 上两个相邻的三联密码子相结合，此时两个 tRNA 分子上 CCA 末端彼此靠近，造成肽链转移的有利的

结构。这个“转动假说”还有待直接的实验证实。

tRNA 结构的研究在过去十几年间是相当活跃的课题，近几年来人们又注意到 tRNA 前体一级结构。tRNA 前体的结构可为 tRNA 生物合成的研究提供重要的信息。tRNA 的转录有三种方式：单顺反子转录、双(多)顺反子转录和含有插入顺序的基因转录。相应地在 tRNA 生物合成中也有三类不同的 tRNA 前体(详见第五章)。

rRNA

rRNA 是核糖体的组成成分(见表 3-4)。它们对于核糖体的自动装配和活力表现起着重要作用。rRNA 结构的研究无疑是了解核糖体结构与功能关系的重要途径之一。近

表 3-4 不同核糖体中的 rRNA*

核糖体	来源	大亚基		小亚基	
		大小	RNA	大小	RNA
80S	真核生物(脊椎动物)	60S	28—29S 5S 5.8S	40S	18S
80S	真核生物(植物和非脊柱动物)	60S	25S 5S 5.8S	40S	16—18S
70S	原核生物	50S	23S 5S	30S	16S
55S (54—61S)	脊柱动物线粒体	40S	16—17S 5S(3S)	30S	12—13S

* 也参阅第六章中表 6-1 和表 6-2。

二十年来 rRNA 结构的研究已经取得了很大的进展，特别是近年来运用快速核苷酸序列测定方法，不少过去难以测定的高分子量的 rRNA 的一级结构得以阐明了。与此同时，综合运用现代生物、化学和物理技术测定了它们的二级结构乃至三级结构，其中以大肠杆菌核糖体中的三个 rRNA——5S rRNA (120 个核苷酸)、16S rRNA(1542)和 23S rRNA (2904) 的结构研究最为详细。其他原核生物及真核生物(包括真核细胞器) rRNA 的研究也取得不少进展，这对于进一步深入研究核糖体的精细结构与功能的关系提供了重要的信息。

小分子 rRNA (如 5S、5.8S)的一级结构可以用经典的或直读的方法进行测定(详见第三节)，而大分子 rRNA 则通常先测定其基因序列并辅之以 RNA 组成和序列分析等手段推导得到。通过电子计算机处理可以排出已知结构的 rRNA 的可能二级结构模型，然后用化学修饰以及酶解敏感部位的确定予以选择和校正；rRNA 上功能部位的研究也有助于其二级结构的建立。

由于 rRNA 分子具柔性，因此企图像 tRNA 那样通过晶体的X射线衍射测定其三级结构是困难的，虽然最近根据小角度X射线衍射分析以及氢动力学研究等已建立了 5S rRNA 的三级结构模型(见图 3-10)。

RNA 是单链结构，它不像 DNA 那样处在完全的螺旋结构之中，但许多分析方法

确认 RNA 中确也存在大量氢键对(见 tRNA 一节的表 3-3)，因此，RNA 分子必然是由茎环结构所构成。tRNA 是这样，rRNA 更是如此。下面将看到，rRNA 特别是大的 rRNA 分子中存在许多由螺旋区和环区组成的区域，每个区域大概是在结构上和功能上相对独立的单位。

5S rRNA

5S rRNA 存在于核糖体大亚基上。真核生物和原核生物的核糖体都有 5S rRNA。从其结构特征不同可以分成四种：(1)真核型(120 个核苷酸，在高等植物中是 116—118 个核苷酸)；(2) 古细菌型(121 个核苷酸)；(3)细菌型(革兰氏阴性菌，120 个核苷酸)；(4)细菌型(革兰氏阳性菌，116 个核苷酸，有的 117 个核苷酸)。

图 3-8 是 Fox 和 Woese 在 1975 年提出的大肠杆菌 5S rRNA 的二级结构模型，当时是根据 6 种原核和 4 种真核生物的 5S rRNA 的一级结构信息提出来的。其后许多实验室提出了许多模型。直至最近 (1982 年)由 De Wachter 等提出的模型，它适用于各种不同来源(包括原核、真核和真核细胞器)的 5S rRNA (见图 3-9)。

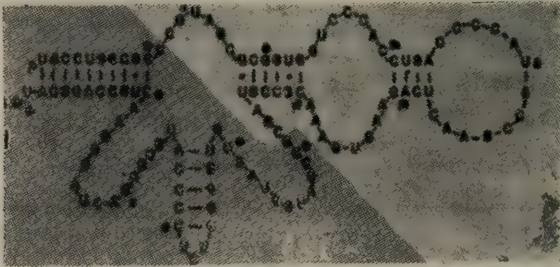


图 3-8 Fox 和 Woese 提出的大肠杆菌 5S rRNA 二级结构模型
图中阴影部分是与 50S 亚基上的蛋白质结合的部分。
从 5′ 端起每 10 个核苷酸用●标记。

根据 Fox 和 Woose 模型，大肠杆菌 5S rRNA 由四个螺旋区及其他单链部分构成。四个螺旋是：(1)分子柄螺旋(1—10 和 110—120)；(2)弱调节螺旋(18—23 和 60—65)；(3)共同臂螺旋(31—34 和 48—51)；(4)原核环螺旋(82—86 和 90—94)。5S rRNA 的分子柄通常由连续伸展的 10—11 对碱基对组成，它把分子的二端维系在一起。不同来源的 5S rRNA 这个区域的核苷酸序列是易变化的，且常常含 G-U 对，例如巨大芽孢杆菌的柄区有三个连续的 G-U 对。

大肠杆菌 5S rRNA 中的弱调节螺旋相当奇特和容易发生变化，这在海生的原核生物中是极弱的，在非海生的原核生物中有中等的稳定性，而在真核生物中则相当稳定。这个螺旋只有六对碱基对，且是高度保守的，其末端的 G-C 对(在大肠杆菌中位于 23 和 60)在所有 5S rRNA 中都是相同的。

由 31—34 和 48—51 构成的螺旋叫共同臂螺旋，共同臂螺旋在所有的 5S rRNA 中都含 4 对碱基对(酵母只有 3 对，是个例外)。紧靠着这个螺旋的环，在绝大多数原核生物中都含有 CGAAC 序列(在大肠杆菌中的 43 至 47 位)，它是 5S rRNA 与 tRNA 分子上的不变序列 GTψCG 相互作用的部位 (见第 69 页)。这种相互作用使 5S rRNA 与 tRNA “打结”，构成一个连续的伸展碱基对，可多至 14 对(如果 5S rRNA 与 tRNA 间有五对碱

基配对的话)，即 5S rRNA 上 4 对，5S rRNA 与 tRNA 间 4—5 对，tRNA 上 5 对。这种分子间螺旋的形成对于 tRNA 上的 TψCG 臂的几何构象，进而对于整个 tRNA 分子三级结构的影响是非常大的。

原核环螺旋一般有 4 对以上碱基对，如含有 G-U 对的话，碱基对可增加，如大肠杆菌则有 8 对。紧靠的原核环(大肠杆菌为 87—89) 有 3 个嘧啶或 4 个非配对的核苷酸。原核环及其螺旋构成一个紧密的发夹。尽管该处一级结构可变，但它总是有较多的 G-C 对，因而比较稳定。大肠杆菌中这个区域是抗核酸酶作用的。

在 5S rRNA 分子上，单链区较双链区的一级结构保守。其中一个单链区，GCGCC$^{G}_{A}$ AUGGUAGU (在大肠杆菌中位于 67—80) 是与 50S 核糖体亚基上的蛋白质如 L5、L11 和 L25 (也见图 3-16)相互作用的部位。这个序列虽普遍存在于细菌 5S rRNA 中，但在真核生物中则找不到。

大肠杆菌 5S rRNA 的第 72 到 83 位，共 12 个核苷酸是与 23S rRNA 上的 5′ 端 143—154 位之间互补的，说明该区域可能是与 23S rRNA 相互作用的部位。

因此，整个 5S rRNA 分子可以粗略地分为二大部分：一部分在识别功能上 (即与 tRNA 相互作用)是重要的，另一部分则在结构上(即与 50S 核糖体亚基相互作用) 是重要的，而隔开这两部分的是弱调节螺旋(见图 3-8)。

1982 年 DeWachter 等人提出一个新的 5S rRNA 二级结构模型，这个模型适用于各种不同来源的近一百种 5S rRNA，表明它们是有共同的起源的。图 3-9 是这个模型的示意图。该模型由以下几部分构成：五个双链区(分别称为 A、B、C、D 和 E)，一个多分支的环(M)，二个中间环(I_1 和 I_2)和二个发夹环 (H_1 和 H_2)。这个模型与上面介绍的大肠杆菌 5S rRNA 模型基本上是类似的。但 DeWachter 等人模型则把所有 5S rRNA 以 150 个位置表示，以容纳在进化过程中核苷酸残基的缺失以及变异产生的核苷酸残基数的增减。此外，还有以下特点：它有较低的自由能，碱基对占碱基总数的 61%；有 21 个位置上的核苷酸残基是保守的，有 24 个位置具有保守的嘧啶或嘌呤碱；而五个双链区的碱基对数目和各个环区的核苷酸数目则因 5S rRNA 的来源不同而不同。

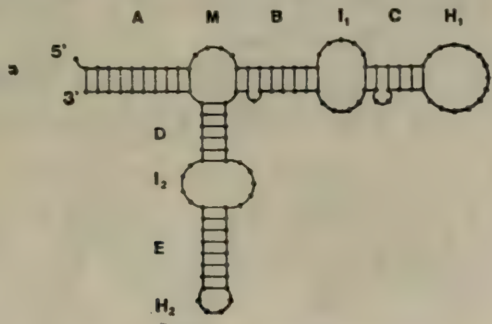


图 3-9 5S rRNA 共同的二级结构模型

A—E：双链区； M：多分枝环；

I_1 和 I_2：中间环； H_1 和 H_2：发夹环。

经过多年努力已有三家实验室分别根据小角 X 射线衍射分析、氢动力学研究和特异性酶解研究等结果提出了大肠杆菌 5S rRNA 的三级结构模型，图 3-10 是其中的一种。其特征是除了二级结构形成的氢键以外(与图 3-8 比较)，还存在着 4 对维持三级结构的氢

键(41—44 和 74—77)，以使 5S rRNA 以更高级结构形式折叠起来(图 3-10)。

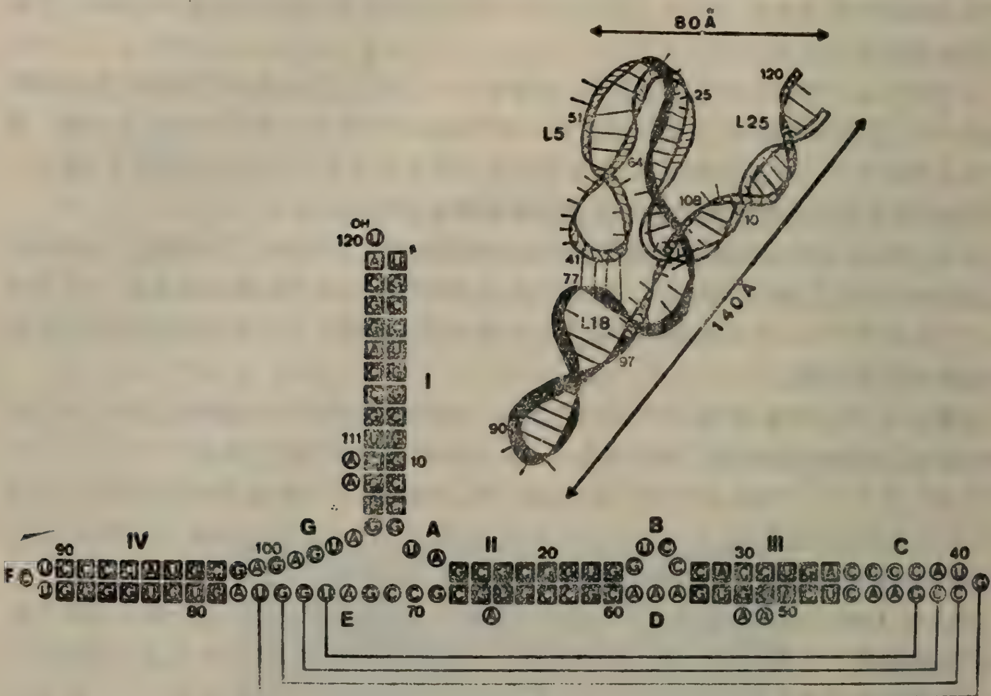


图 3-10　大肠杆菌 5S rRNA 的三级结构模型(与图 3-8 比较)

5.8 S rRNA

真核生物核糖体大亚基 RNA 的组成与原核生物的最明显的差别是含有 5.8S rRNA。5.8S rRNA 是在变性条件下从 28S rRNA 上解离出来的（见第五章）。已有十多种 5.8S rRNA 的一级结构被测定了。与 5S rRNA 相比较，5.8S rRNA 略大些，通常含有 160 个左右的核苷酸残基；分子中含有修饰核苷酸(如 ψ 和核糖上甲基化的核苷酸等)；其一级结构中含有原核生物 5S rRNA 的共同序列 CGAAC（见图 3-11）。可见它可能是真核生物蛋白质生物合成时参与和 tRNA 相互作用的部位，起着原核生物 5S rRNA 相似的功能。不同来源的真核 5.8S rRNA 之间的共同性约为 75%。5.8S rRNA 的二级结构已有

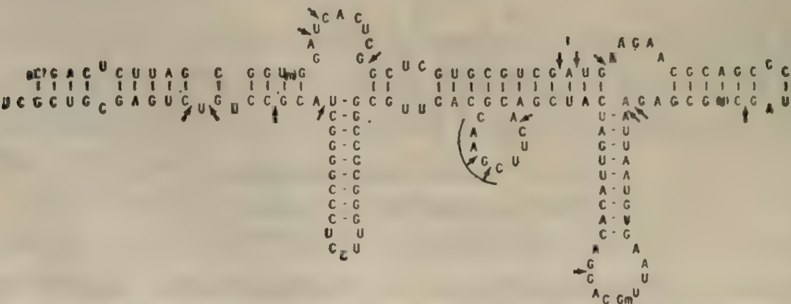


图 3-11　Novikoff 腹水瘤 5.8S rRNA 可能的二级结构模型

图中特别标出的$^{5'}$CGAAC$^{3'}$(103—107) 是可能与 tRNA 中$^{5'}$GTψCG$^{3'}$ 相互作用序列

许多人提出了许多可能的模型，但迄今没有定论。图 3-11 显示的是一种可能的模型，这个模型的特点是含有一个富有 G-C 碱基对的发夹环（116—138），一个富有 A-U 碱基对的稳定性较差的发夹环(64—91)和二个近似对称的突起（分别在 15—25 和 40—44）。

16S rRNA

已有几种原核生物(或真核生物细胞器)的 16S rRNA 的一级结构被阐明了，它们是大肠杆菌 16S rRNA（1542 个核苷酸），普通变形杆菌（*Proteus vulgaris*）16S rRNA(1544 个核苷酸)，小眼虫（*Euglena gracilis*）叶绿体 16S rRNA（1491 个核苷酸)和衣藻（*Chlamydomonas reinhardi*）叶绿体 16S rRNA（1475 个核苷酸)等。这些 16S rRNA 中除四种常见核苷酸外，还含有少量修饰核苷酸，如在大肠杆菌 16S rRNA 中有五种——m^6A、m^2G、m^5C、m^4C 和 Gm。不同来源的 16S rRNA 具有相似的二级结构；并且发现核糖体小亚基 rRNA（包括真核 18S rRNA，线粒体 12S rRNA）的二级结构模式也是类似的（见后面图 3-13 及正文说明）。下面以大肠杆菌 16S rRNA 为例简单介绍 16S rRNA 的结构特征。图 3-12 是 Noller 和 Woese 提出的大肠杆菌 16S rRNA 的二级结构模型。

该模型有下列特征：

整个分子被三个分子内的长距离相互作用形成的螺旋分隔成 4 个相对独立的部分，这就是 5′ 末端区域（I，1—556)、中心区域（II，564—916)、3′ 末端大区域(III_1，926—1391) 和 3′ 末端小区域（III_2，1392—1542)。在该模型中约有 46% 的碱基处在螺旋结构之中(这与 tRNA 有 50%、5S rRNA 有 45%的双链区数值是接近的)，但用物理方法测定则有 60—70% 双链结构。这个差异表明，在 16S rRNA 中肯定也存在三级结构，二级结构中不配对区域之间可能包括非 Watson-Crick 的相互作用。另有证据说明，5′ 末端区域和 3′末端大区域中的不配对环至少是部分结构化了的。

在大肠杆菌 16S rRNA 中，许多螺旋含有 G-U 对，有的（例如 830—860）含有多个 G-U 对。分子中不存在长的、连续的标准氢键对，最长的螺旋只有 8 对。这与 16S rRNA 前体中处于成熟 16S rRNA 旁侧的 RNase III 剪切部位有很大区别，该茎区由 17 个连续的标准碱基对构成。在 16S rRNA 分子的一些茎区中还出现单碱基突起，如 G_{31}、A_{746}、A_{1227} 和 A_{1441}，至少其中二个(G_{31} 和 A_{746}) 被认为与核糖体蛋白质的识别有关。

与 tRNA 一样，16S rRNA 也存在着复合螺旋结构，几个(多至 5—6 个)短的螺旋处在一个复合螺旋之中，它们被环所隔开。

在 16S rRNA 中有几处的核苷酸序列能与多于一处的分子的其余部分互补，这种多相互作用结构被称作“开关”结构，例如在 3′末端大区域的大环区就存在这种结构。“开关”结构的存在是 16S rRNA 在核糖体中进行蛋白质生物合成的不同阶段上发生构象变化的基础。但是，统观整个 16S rRNA 分子，其中并不存在“打结”的结构，即不存在这样的区域：一个碱基配对区的一条链处于另一螺旋的两条链之间，也即邻近螺旋的环区并不与另一条链互补。这可能对核糖体的装配有重要意义。因为没有“结”的结构可以避免在核糖体装配时产生中间物——同分异构象物（conformers)。实际上，已知 RNA（包括 5S rRNA、tRNA 和 16S rRNA）的结构都是按照这个原则构成的。

假定核苷酸间的平均距离为 3 Å，那么图 3-12 的 16S rRNA 结构大小为 300 × 450 Å。从小角度 X 射线衍射和电镜观察到，其大小分别为 55 × 200 × 220 Å和 80 × 190 ×

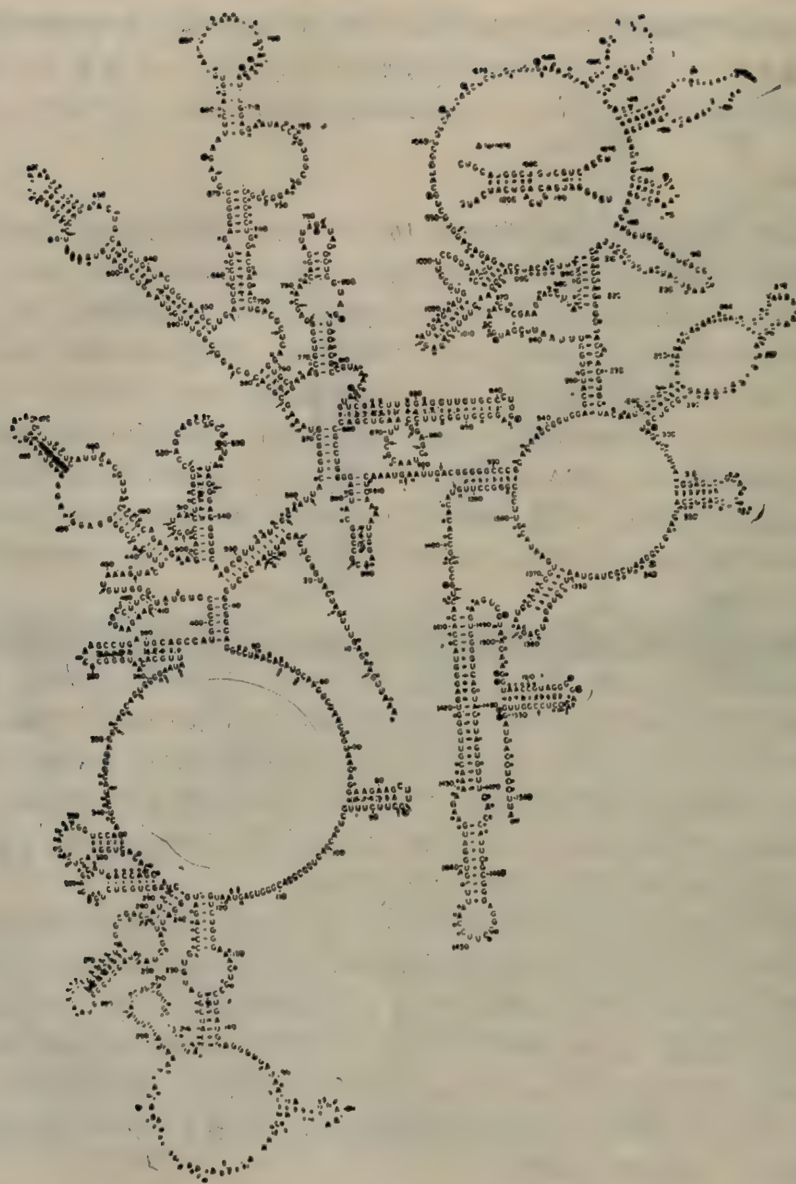
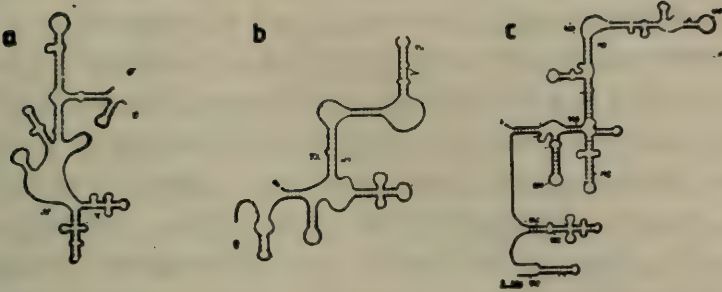


图 3-12 大肠杆菌 16S rRNA 的二级结构模型

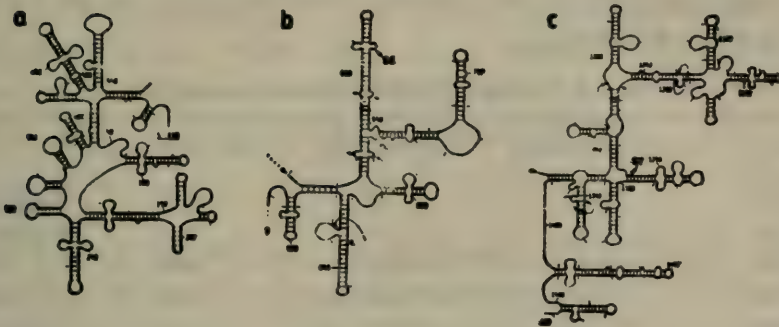
250 Å。因此，16S rRNA 是被压缩在核糖体 30S 亚基内的。

rRNA 基因是细胞内所有基因中最保守的。通过测定三个主要种系(包括原核生物、真核生物和高等动植物细胞器的)核糖体的小亚基 rRNA 的核苷酸排列顺序，得到的结果是它们有广泛的共同性，例如大肠杆菌 16S rRNA (如前述)，酵母 18S rRNA (1,789 个核苷酸)和人线粒体 12S rRNA (954 个核苷酸)。综合这些结果，说明核糖体小亚基上的 RNA 在进化上是保守的，特别是它们的二级结构显得更为保守，见图 3-13。这种广泛的保守性意味着 rRNA 有着很强的功能上的制约。从多种来源的 16S rRNA 的 RNase T_1 酶解图谱看到，有一些序列在进化过程中是始终不变的，而这些保守区一般处于分子的非

人线粒体 12S rRNA



大肠杆菌 16S rRNA



酵母 18S rRNA

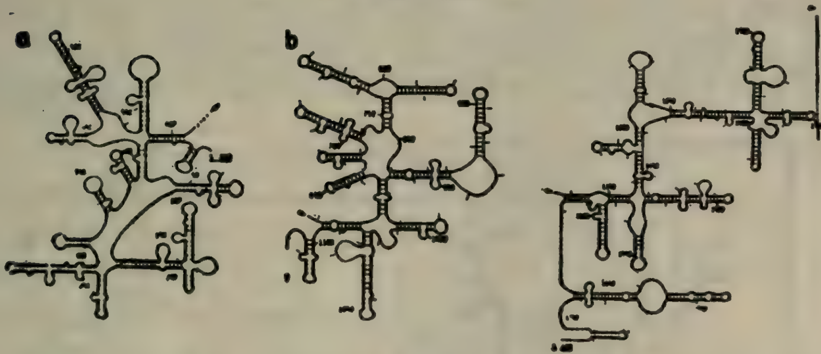


图 3-13 核糖体小亚基 RNA 二级结构的比较

a，代表 5′端区域； b，中心区域； c，3′ 端区域(包括大区域和小区域)。

螺旋区。在螺旋区中，核苷酸的变化是相互制约的，即其中一条链改变，其对应的链也在变，即使同一种族也是如此。这些螺旋往往含有高比例的非标准碱基对如 G-U 甚至 A-G 对。不同来源的 16S rRNA 分子中的这些区域在长度上是可变的，它们中的一部分是与蛋白质相结合的部位。

研究核糖体上核酸与蛋白质相互作用的方法有下述 12 种：(1)蛋白和核酸结合后核酸酶水解的研究；(2)核酸片段与个别蛋白质结合能力的研究；(3)蛋白质肽链被裂解成二个或更多的片段，然后分别测定与 16S rRNA 结合的能力；(4)先经核酸酶作用，然后分

离和分析得到的 RNA-蛋白质络合物中的 RNA 片段和所含的蛋白质；(5)交链分析，即先经紫外光照射或双功能试剂处理使分子间发生交链，然后确定彼此交链的蛋白质和 RNA 片段；(6)单链特异性的试剂处理，例如，在一个蛋白质结合到 RNA 链上前后分别用乙氧羰基丁酮醛处理，分析两者被修饰部位的差异，从而判断与蛋白质作用的部位；(7)蛋白质分子上特异性的氨基酸被修饰后，对 RNA 结合能力的影响；(8)突变型的蛋白质与野生型的蛋白质结合 RNA 能力的比较；(9)电子显微镜观察 RNA 与蛋白质结合位点；(10)用超滤膜和其他方法测定 RNA 与蛋白质的结合常数；(11)物理技术、例如小角 X 射线衍射、中子衍射、氢动力学分析和圆二色性测量等等，能够推测出蛋白质结合到 RNA 链上以后引起 RNA 分子构象的变化；(12)天然核糖体蛋白质的萤光分光光度研究，能给出蛋白质结合后有关氨基酸微环境的变化的信息。综合以上这些研究得到的结果见表 3-5 和图 3-14。从中可以看出，16S rRNA 链上的这些结合位点在结构上是多种多样的，在长度上也是长短不一的，如 S8 和 S15 的结合部位约为 40—60 个核苷酸残基而 S4 则大约有 500 个核苷酸残基。

表 3-5　大肠杆菌 16S rRNA 的结构部位与功能的关系

结　构	16 SrRNA 上的部位	结合的蛋白质	在核糖体上的定位	功　能
5′ 末端区域	25—420	S4、S20	基部	与核糖体上蛋白质结合
5′ 末端区域	420—460	S4	基部	与核糖体上蛋白质结合
中心区域	560—890	S6、S8、S15、S18	平台	亚基结合
3′ 末端大区域	920—1400	S7、S9、S10、S13、S19	头部	与 tRNA 结合
3′ 末端小区域	1400—1541	S1 和 S2	平台	亚基结合、起始

除与核糖体上蛋白质相互作用外，16S rRNA 还与参与蛋白质合成有关的核糖核酸相互作用。16S rRNA 的 3′ 末端区域存在着一段序列 ACCUCCU 是与 mRNA 的翻译起始区相互补的(见第六章)，这一区域处于核糖体 30S 亚基的平台上或在裂缝中，可能裂

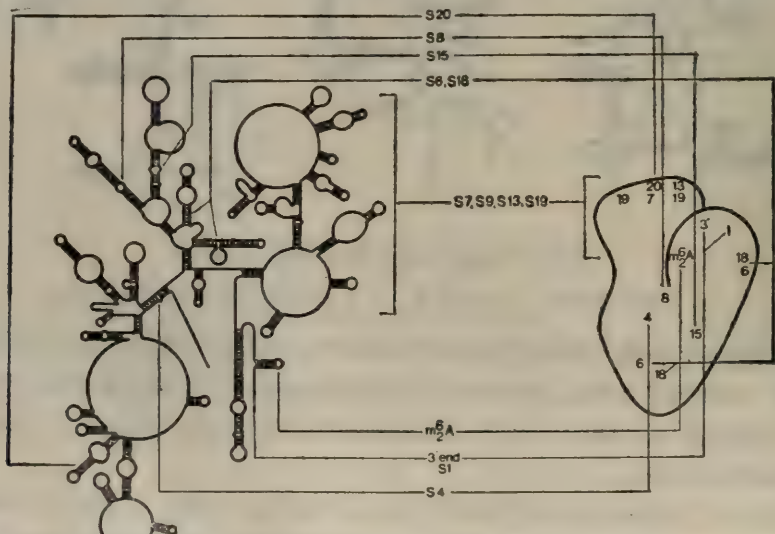


图 3-14　大肠杆菌 16S rRNA 二级结构在核糖体 30S 亚基上的定位
(亚基模型见第六章)

缝区本身形成一个与 mRNA 结合和密码子与反密码子辨认的部位。3′ 末端小区域中二个不配对的区域有高度的保守性，已有许多证据说明它们与核糖体 50S 亚基结合有关。

有的 tRNA 占据 P 位时，分子上的反密码子的 5′末端核苷酸能与 16S rRNA 通过紫外光照射高效地共价交链在一起。这个交链部位在 16S rRNA 分子的 1390—1497，1390—1410 是十分保守的，有可能这就是各种核糖体小亚基 rRNA 起关键的和共同的功能的部位。这样，密码子和反密码子相互作用点被置于 3′末端大小区域之间的连接部位，而这个位置有人推测正好在 30S 亚基电子显微镜模型的头部和平台的连接部位，这就是有可能在该模型的裂缝处，这一点与核糖体上和 tRNA 有很强的结合力的蛋白质在该模型上的位置是一致的(就是 S2、S3、S10、S1 4、S18、S19 和 S21)，这些蛋白质都环绕着裂缝区，mRNA 的结合部位也在该处。

核糖体 30S 亚基和 50S 亚基的结合也有 16S rRNA 参与。核酸酶限制性水解和化学修饰研究显示，当与 50S 亚基结合以后，其中心区域的 3′末端大区域和 3′末端小区域是绝对地抗核酸酶和 β-乙氧羰基丁酮醛作用的。 这些结合部位上存在着起关键作用的 G 残基。由此看来 16S rRNA 的中心区域、3′ 末端大区域和小区域在核糖体行使生物功能上起着重要作用。这些区域的单链区的结构显示了种属间的保守性。

18S rRNA

迄今只有二种真核生物 18S rRNA 的一级结构被确定，它们是通过测定基因的一级结构并辅之以 RNA 组成和部分 RNA 序列分析结果推导出来的。前面已经提到，酵母 18S rRNA 含有 1789 个核苷酸，其中有约 25 个甲基化核苷酸如 amϕ 即 $m^1ncp^3\phi$、m^7G、m^6A、m^6_2A 和各种 2′-0-甲基核苷酸和许多假尿苷酸；爪蟾 18S rRNA 含有 1,825 个核苷酸，其中有 40 个甲基化核苷酸和许多假尿苷酸。这两种 rRNA 的核苷酸序列有 70% 以上是相同的，也存在部分相同以及完全不同的区域。几乎所有甲基化的修饰成分都处在结构高度保守的区域。

对鸡、鼠 18S rRNA 的一级结构也进行了测定，发现它们与酵母和爪蟾有许多相似性，特别是含有修饰成分的片段。例如鼠 18S rRNA 中的 UAm^{6}ACAAG 和 $m^6_2Am^6_2$ACCUG 片段分别存在于酵母和爪蟾的靠近 3′末端区域；鼠中有 m^7GAAU (ϕ, C_3) AG，而酵母和爪蟾中则分别是 GAAUUCCUAG 和 m^7GAAUUCCCAG。此外，和爪蟾相似鼠 18S rRNA 中的修饰成分比酵母多。

虽然真核生物 18S rRNA 间有很强的保守性，但与原核生物 16S rRNA 相比，除了 3′ 末端外，一级结构很少有共同序列。图 3-13 比较了三种不同来源(原核、真核和真核细胞器)的核糖体小亚基 rRNA 的二级结构，发现它们是非常保守的，即分开为几个区域都具有相似的二级结构。 看来二级结构上的保守区域可能是这类 rRNA 行使功能的关键部位。

23S rRNA

1980 年 Brosius 等利用 23S rRNA 基因顺序测定方法推导出大肠杆菌 23S rRNA 的全部核苷酸排列顺序，整个分子含有 2904 个核苷酸。已经测定了一级结构的还有玉米叶绿体 23S rRNA。1981 年有三家实验室相继提出了大肠杆菌 23S rRNA 的二级结构模

型，其中 Noller 等人提出的模型见图 3-15。整个分子被许多长距离相互作用的碱基对分成相对独立的六个部分，其中 5′末端与 3′末端的一段序列（8 核苷酸）互补，构成一个闭合环。中心环的四周呈放射状。该模型中约有 52% 的核苷酸残基处于螺旋结构之中，螺旋区是不规则的，大体有下列三种情况：（1）由标准 Watson-Crick 碱基对构成的螺旋；（2）由许多 G-U 对组成的螺旋；（3）含有特殊的碱基对（如 A-G 对）和包含有碱基突起的螺旋。

从图 3-15 可以看到，像 16S rRNA 一样，23S rRNA 也不存在"打结"结构，但含有"开关"结构，这些结构特征暗示或许它们在功能上是重要的。

23S rRNA 分子上存在与 tRNA、5S rRNA、5.8S rRNA 以及叶绿体 4.5S rRNA、昆虫 2S rRNA 相互作用的位点。如 23S rRNA 有一段顺序（1984—2001）与同源的 tRNA$_f^{Met}$ 上的一段顺序（G_5-D_{21}）互补；靠近 5′末端的一段 12 核苷酸（143—154）与 5S

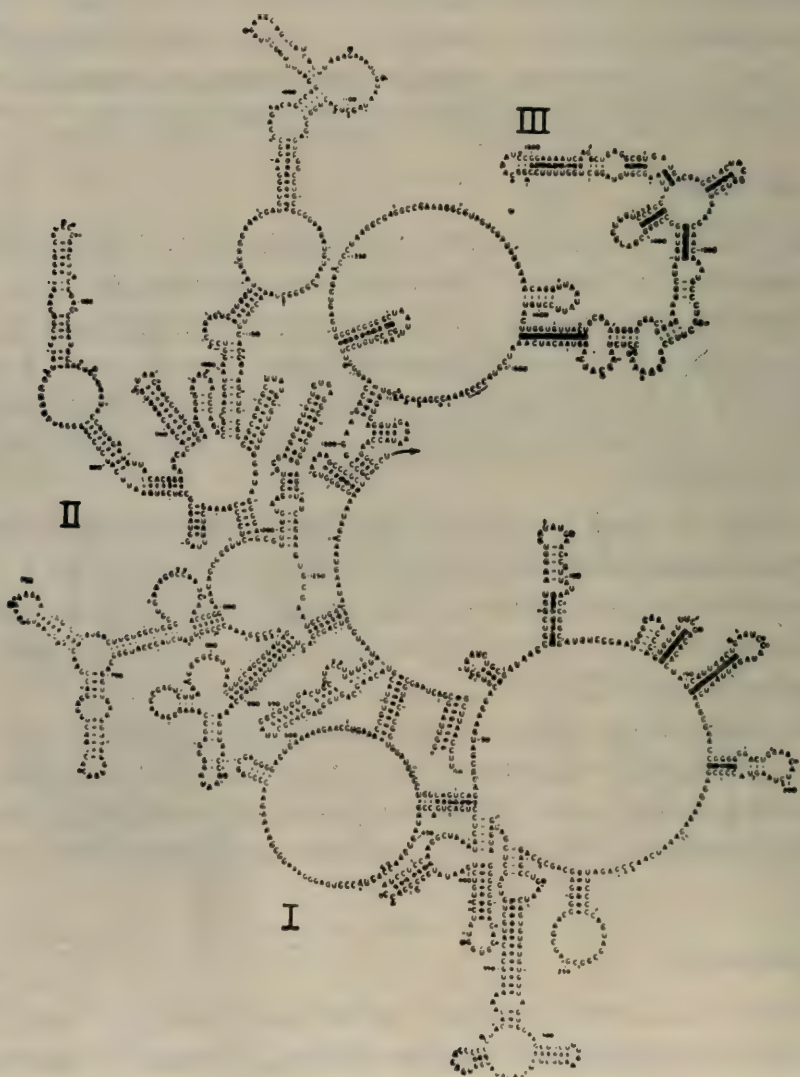
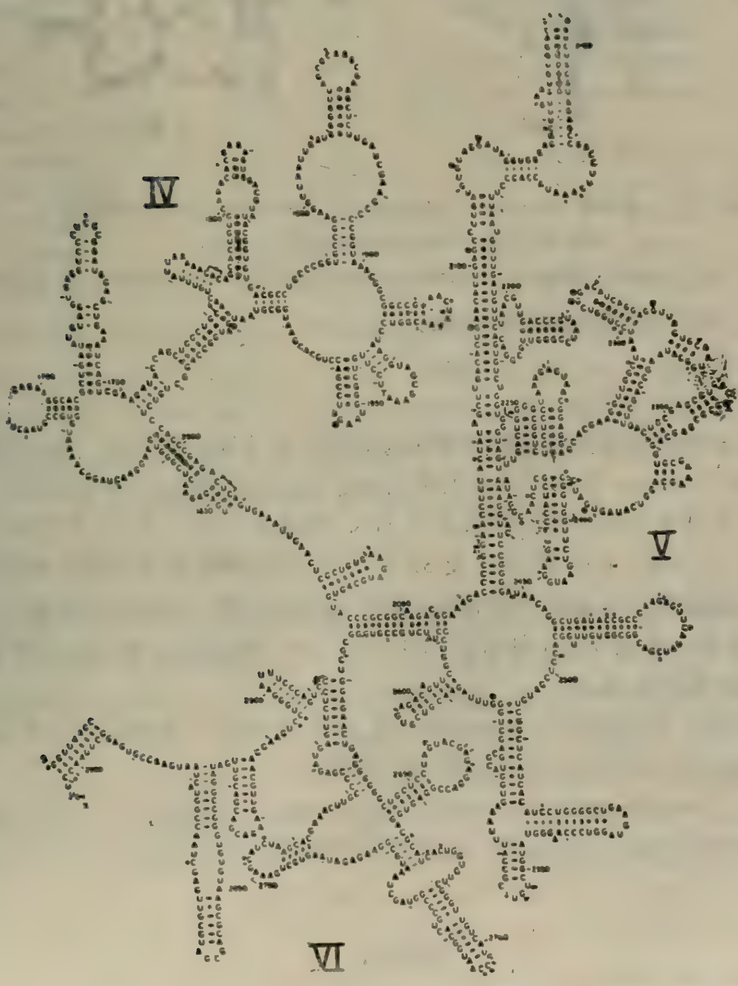


图 3-15　大肠杆菌 23S rRNA 的可能二级结构[左半有

rRNA 上一段顺序(72—83)互补;5′末端区域约有 160 核苷酸对应于不同真核生物的 5.8S rRNA 顺序; 3′末端有一段顺序则对应于叶绿体核糖体大亚基所含有的 4.5S rRNA; 5′末端区域有 20 多个核苷酸对应于某些昆虫 80S 核糖体上的 2S rRNA 等。 这些对应关系意味着它们之间有功能上的联系。

嘌呤霉素亲和标记核糖体时,其作用位点在 A 位上(见第六章), 该化合物能与 23S rRNA 上的 GU*CCG 或 GU*UCG 相互作用,前者位于 23S rRNA 上的 962—966,出现一次,后者位于 2553—2557, 2603—2607 和 2688—2692, 出现三次。如果 23S rRNA 在核糖体行使功能时确实发生这些相互作用,那么这些区域是重要的。

约有 20 种蛋白质, 即核糖体大亚基 50%以上的蛋白质能与 23S rRNA 不同程度地结合。根据抗核酸酶作用、紫外光照射和双功能试剂交链的研究, 指出 23S rRNA 上第 615 位的 U 与 L4 蛋白质上第 35 位的 Tyr 相交联。其他可能结合部位见图 3-16, 其中特



三个区域(1—1647 核苷酸),右半也有三个区域(1648—2904)]

别是 L1 蛋白质的结合区，经过比较几种不同来源 23S rRNA 在该处的一级结构，说明进化上是保守的。

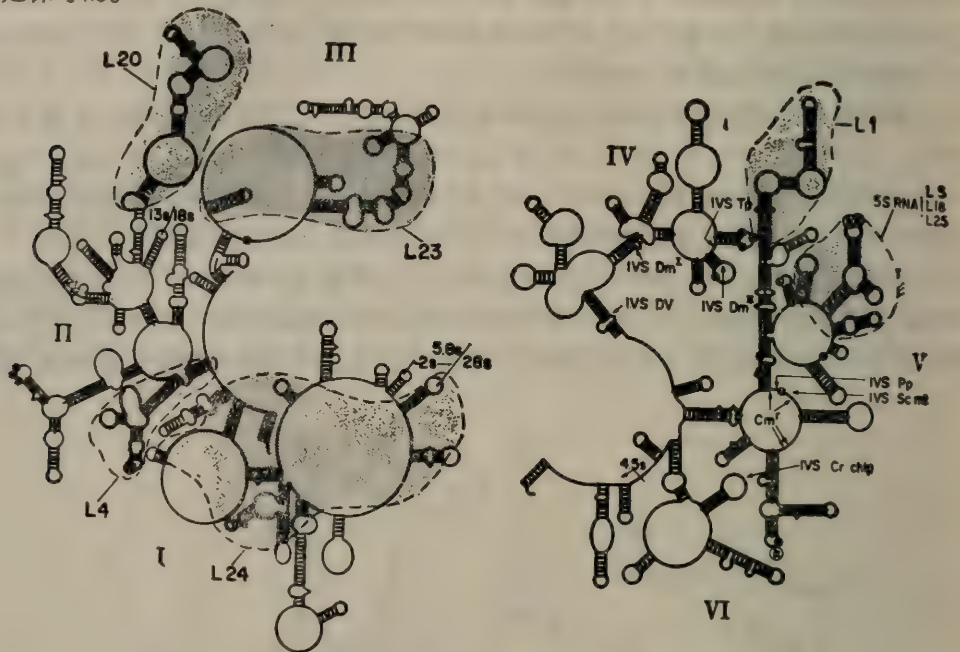


图 3-16 大肠杆菌 23S rRNA 二级结构模型上 23S rRNA 的可能功能部位
（图中阴影处系与其他生物大分子作用的部位）

不同来源的 23S rRNA 约有 60—70% 的结构共同性。大肠杆菌 23S rRNA 的一级结构与酵母 21S rRNA 也有很大的相似性。

26S rRNA

已经测定了一级结构的真核生物核糖体大亚基中大 RNA 的只有一种，就是酵母 26 S rRNA，它是由 Veldman 等人运用基因序列测定的结果并辅之以 RNA 分析结果推导出来的。整个分子含有 3393 个核苷酸；其中 G-C 含量占 48%；有 43 个甲基化的修饰成分（如 m^1A、m^1G、m^5C、m^3U 和 m^5U 及 37 个 2′-O-甲基核苷），还有 36—37 个假尿苷。在 26S rRNA 上修饰成分不是随机分布的，常常成串地出现，而在分子的某些区域则几乎没有修饰成分，如 5′末端 600 个核苷酸和 3′ 末端 350 个核苷酸。修饰成分多出现在 26 S rRNA 上的进化保守区域。

酵母 26S rRNA 与爪蟾 28S rRNA 相比较，在 5′ 末端和 3′ 末端的序列相同性较高，而与大肠杆菌 23S rRNA 相比，只有少数序列相似性。

和其它大分子 rRNA 相似，26S rRNA 整个分子被许多长距离相互作用的碱基对构成的螺旋分隔成若干区域，26S rRNA 共分 7 个区域，这些区域可能构成分子的功能部位。

mRNA

1961 年 Jacob 和 Monod 提出用信使 RNA（mRNA）这个名称来命名带有从 DNA

上得到信息并指导蛋白质合成的一类 RNA 分子。mRNA 占细胞 RNA 总量的 3—5%。mRNA 的平均分子量约 500,000,沉降系数为 8S,但是不同 mRNA 在长度和分子量上有很大的不同。在大肠杆菌中,mRNA 的平均长度约 900 到 1,500 个核苷酸残基,可以编码 300 到 500 个氨基酸。部分真核 mRNA 的长度见表 3-6,其中大的 mRNA 如丝心蛋白 mRNA 有 19,000 个核苷酸。

表 3-6 部分真核 mRNA 和它们的聚腺苷酸的长度

mRNA	整个分子大小(核苷酸数)	编码区大小(核苷酸数)	测长度方法	聚腺苷酸长度
鼠胰岛素	354	258	序列分析	—
兔珠蛋白	(α)551	423	序列分析	30—40
	(β)589	438	序列分析	
人 β 珠蛋白	626	444	序列分析	—
小鼠珠蛋白	703	463(平均)	凝胶电泳	40、60、100
小鼠免疫球蛋白轻链	1,200	660	凝胶电泳	200
肌球蛋白重链	6,500	5370	凝胶电泳	170
牛眼 α-A_2 晶体	1,460	520	凝胶电泳	200
蚕丝心蛋白	19,000	16,000	凝胶电泳	100
海胆组蛋白 H_4	400—420	306	凝胶电泳和序列分析	无
两栖类动物卵母细胞组蛋白 H_4	—	—		有
小鼠胰 α-淀粉酶	1,577	1524	DNA 序列分析	—
小鼠唾液腺 α-淀粉酶	1,659	1533	序列分析	—
小鼠肝 α-淀粉酶	1,773(主)	—	序列分析	—
小鼠肝 α-淀粉酶	1,806(次)	—	序列分析	—

原核生物和真核生物的 mRNA 的结构是不同的。动植物病毒 RNA 既是遗传信息的载体也起 mRNA 的作用,这些 RNA 也与真核生物 mRNA 一样是单顺反子结构的,不过它先被翻译成一个大的蛋白质,然后再被水解成有生物活性的肽链。下面简要介绍原核和真核生物 mRNA 及病毒 RNA 的结构。类病毒虽然不起 mRNA 的作用,因其与病毒 RNA 一样有侵染性,故在此一并讨论。

原核生物 mRNA 的结构

原核生物 mRNA 在结构上十分相似于噬菌体 R17、MS2、Qβ 和 f2RNA。原核生物 mRNA 的典型结构是多顺反子的,即一个操纵子负责的一段转录产物对几条多肽链编码,5′端无帽子结构,分子中一般没有修饰核苷酸。众所周知的例子是大肠杆菌乳糖操纵子 mRNA,编码三条多肽链;色氨酸操纵子编码五条多肽链。也有单顺反子的细菌 mRNA,例如大肠杆菌脂蛋白 mRNA。而真核生物 mRNA 一般是单顺反子的。下面以噬菌体 MS2RNA 为例说明原核生物 mRNA 的结构特点。1976 年 Fiers 等测定了噬菌体 MS2 RNA 全部 3569 个核苷酸的排列顺序,它由三个顺反子组成(以后又修正为四个顺反子)(见图 3-17)。从图 3-17 可见整个 MS2 RNA 是一条单链。其 5′端为鸟苷三磷酸,由 5′端向 3′端顺次是 5′端非编码区(129 个核苷酸),A 蛋白(成熟蛋白)基因(1179 个核苷酸),终止密码子和间隔区(26 个核苷酸),外壳蛋白基因(390 个核苷酸),溶菌蛋白基因(225 个核苷酸),复制酶基因(1635 个核苷酸)及终止密码子和 3′端非编码区(174 个

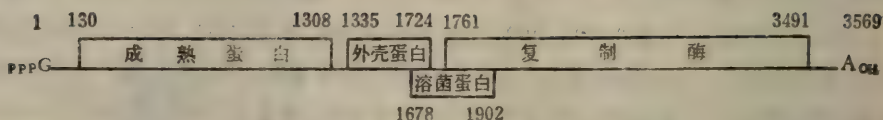


图 3-17 噬菌体 MS2 RNA 的结构示意图

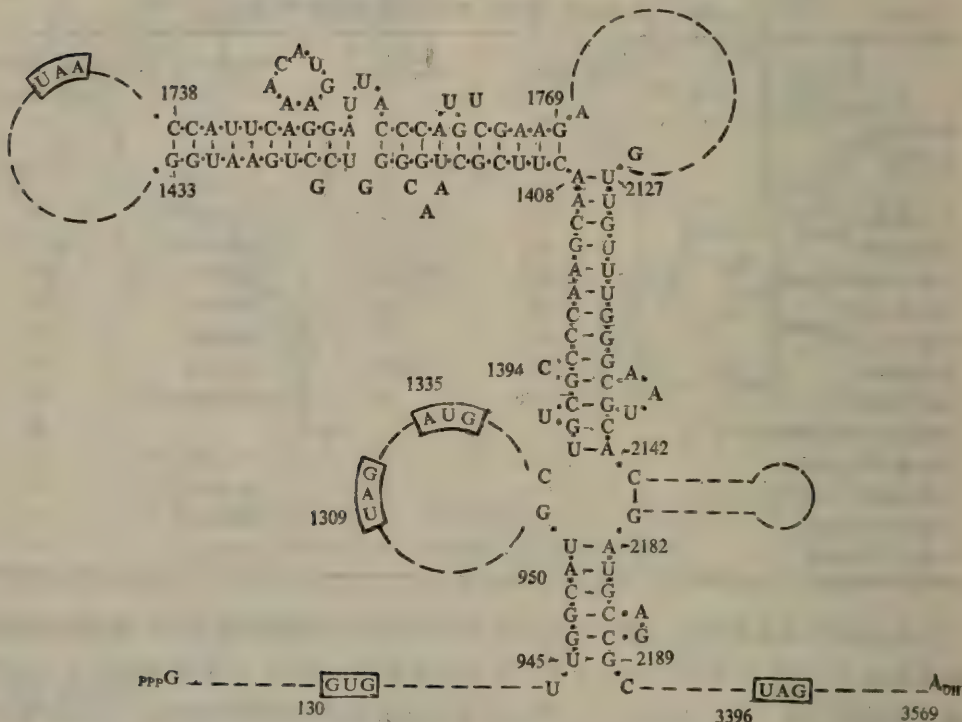


图 3-18 噬菌体 MS2 RNA 分子内长距离相互作用模型

核苷酸)。其中溶菌蛋白基因中有 189 个核苷酸与相邻两个基因共用。整个分子中无修饰该苷酸。Fiers 等人报告 MS2 RNA 三个基因的一级结构的同时也提出了它们的二级结构(见图 3-18 和 3-19)。图 3-18 和 3-19 分别是 MS2 RNA 的长距离相互作用和外壳蛋白基因的花束状二级结构图。在 MS2 RNA 分子上存在着丰富的自身回折所产生的双链区，几个起始密码子都处于分子的单链区域。这个二级结构图主要是根据 Watson-Crick 碱基配对原则和核酸酶部分酶解的数据确定的。

真核生物 mRNA 的结构

图 3-20 是真核生物 mRNA 一级结构的通式的示意图，整个分子由 5′ 端帽子结构，5′端非编码区(不翻译区)，编码区(翻译区)，3′端非编码区和 3′端聚腺苷酸组成。

与噬菌体 RNA 一样，真核 mRNA 和高等动植物病毒 RNA 也具有丰富的二级结构。如鸭珠蛋白 mRNA 用溴乙锭结合光学性质的测定指出，有 45—60%的核苷酸处于双螺旋之中；兔珠蛋白 mRNA 热变性的实验指出有 55—62%的核苷酸处在碱基配对之中。

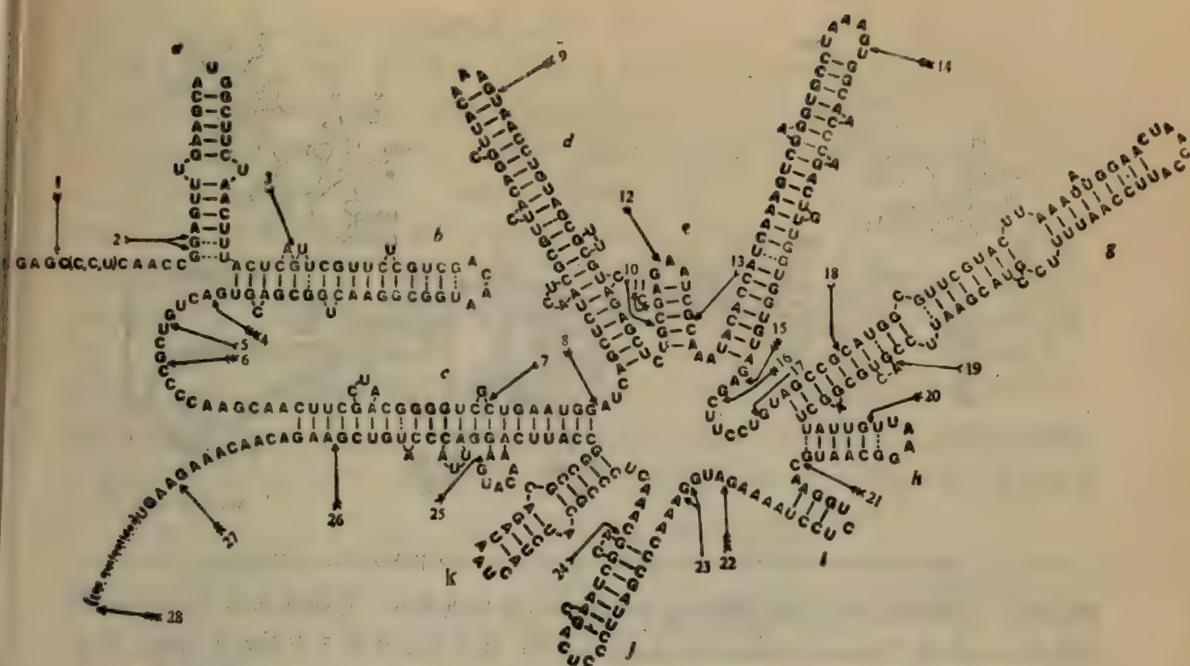


图 3-19 噬菌体 MS2 RNA 分子中外壳蛋白基因的二级结构图

(↑表示 RNase T_1 限制性水解点)

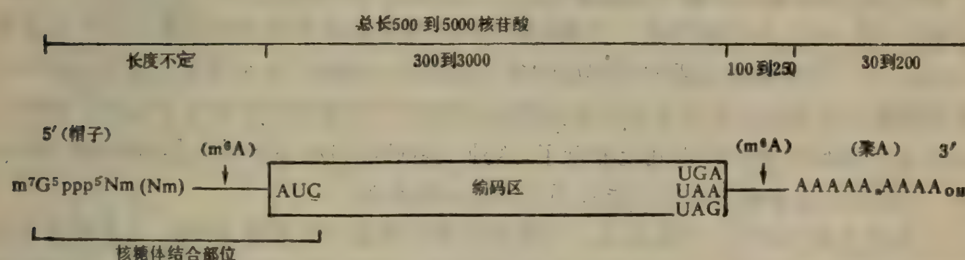


图 3-20 真核生物 mRNA 的结构示意图

近年来由于电子计算机在结构分析中的应用,已推测出 mRNA 确有丰富的二级结构(见图 3-21)。

原核生物和真核生物 mRNA 结构的比较

(1) 真核生物 mRNA 有帽子结构,原核生物 mRNA 无帽子结构。真核生物 mRNA 5′端帽子结构见表 3-7 和图 3-22。

帽子结构通常有三种类型 ($m^7G^{5'}ppp^{5'}Np$, $m^7G^{5'}ppp^{5'}NmpNp$ 和 $m^7G^{5'}ppp^{5'}NmpNmpNp$),详见表 3-7。除真核生物 mRNA 外,5′端帽子结构也是高等动植物病毒 mRNA 所特有的,有的病毒 RNA 具有特殊的帽子结构(详见后述)。这些帽子结构都是转录后早期加上去的(见第五章)。

帽子结构对于 mRNA 的翻译活性是重要的。它可使 mRNA 免遭外切核酸酶的降解

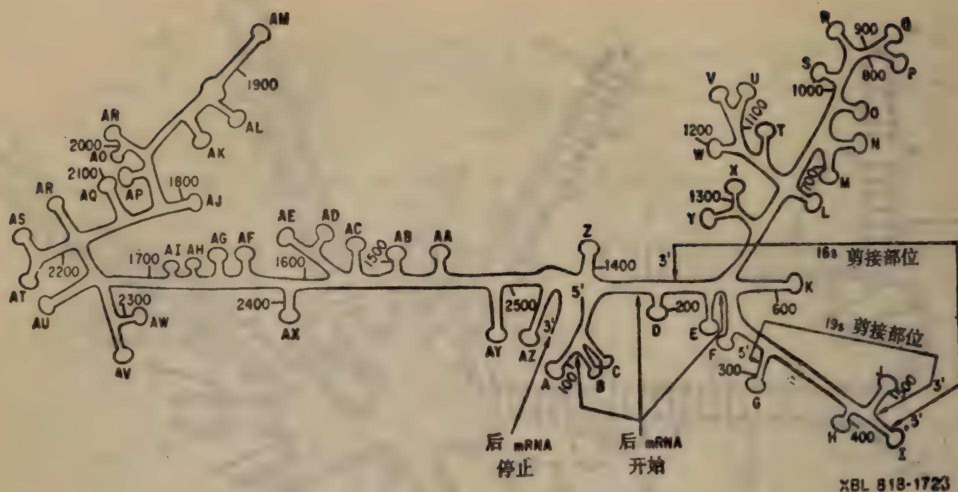


图 3-21 SV40 病毒后 mRNA 前体的二级结构模型(英文字母表示 52 个环区部位)

和 mRNA 与核糖体的结合包括翻译起始区的辨认都是重要的。呼肠孤病毒 (reovirus)和水泡性口炎病毒 (vesicular stomatitis virus) mRNA 都有甲基化的 5′末端帽子结构，带 5′端甲基化的 mRNA 比 5′端不带甲基化的在麦胚系统中合成蛋白质的速度快得多。如果用化学方法除去 m^7G，病毒 RNA 和兔珠蛋白 mRNA 的绝大部分模板活性就消失了。

5′端帽子结构与蛋白质翻译的起始有关。无 m^7G 的呼肠孤病毒不能与 40S 亚基形成复合物，而有 m^7G 的则能形成。合成的有 m^7G 的短的寡聚核苷酸也能与 40S 亚基形成复合物。m^7G 可能是合成蛋白质起始的辨认信号的一部分，如小鸡卵白蛋白 mRNA，其5′末端的帽子结构附近区域能形成发夹结构，见图 3-23。这样的结构使 5′端帽子结构与蛋白质合成的起始密码子在空间上靠近了，这种 5′端附近的发夹结构在其他真核 mRNA 上也有，因此有人推测帽子结构与蛋白质合成的启动有关。

(2) 真核生物 mRNA 一般都有 3′端聚腺苷酸。见表 3-6 和图 3-20。它是转录后早期由腺苷酸聚合酶催化加上的(见第五章)。其长度一般为 20—200 个核苷酸，早期的研究发现，组蛋白 mRNA 和真核生物病毒 RNA 无聚腺苷酸，但近年来的研究已证实，组蛋白 mRNA 也有聚腺苷酸结构，如两栖类卵母细胞组蛋白 H5 的 mRNA 有聚腺苷酸（表

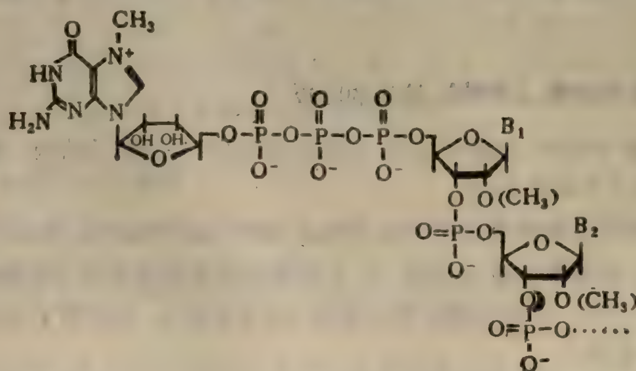


图 3-22 真核生物 mRNA 5′端帽子结构通式

3-6)；植物病毒如大麦条纹花叶病毒 mRNA 等也发现有 3′ 端聚腺苷酸。原核生物 mRNA 一般没有或仅有很短（少于 10 个）聚腺苷酸结构。目前一般认为聚腺苷酸对于 mRNA 的翻译功能并非必须，如小鼠 L 细胞 mRNA，除去其 3′端聚腺苷酸并不影响它的翻译活性。另外，珠蛋白 mRNA 和除去聚腺苷酸的珠蛋白 mRNA，两者起初有相同的翻译活性，但是全部除去聚腺苷酸的 mRNA 和少于 20 个腺苷酸残基的 mRNA 其活性很快就降低了。这些实验表明 mRNA 的降解与聚腺苷酸的存在与否以及其长度有关。但是聚腺苷酸是如何影响 mRNA 降解的目前还不清楚。此外，聚腺苷酸还与 mRNA 顺利通过核膜进入细胞浆的过程有关；还有人认为聚腺苷酸与病毒的侵染性有关。

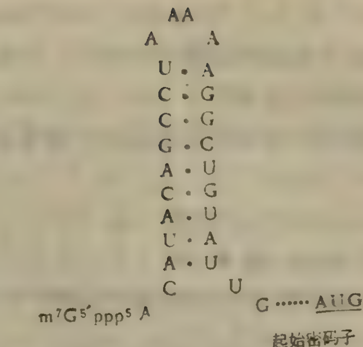


图 3-23　真核 mRNA 5′端附近的发夹结构

表 3-7　细胞内 mRNA 5′端结构和甲基化核苷

mRNA	5′ 末端结构	分子中间甲基化核苷
HeLa 细胞	$m^7G^{5'}ppp^{5'}N'mpNp$, $m^7G^{5'}ppp^{5'}N'mpN''mp$ ($N'm = m^6Am, Am, Gm, Um, Cm$; $N''m = Gm, Um, Cm$)	m^6A
小鼠骨髓瘤	$m^7G^{5'}ppp^{5'}NmpNp$, $m^7G^{5'}ppp^{5'}NmpNmp$	m^6A($1.8m^6A:1m^7G$)
大鼠肝细胞瘤	$m^7G^{5'}ppp^{5'}NmpNp$, $m^7G^{5'}ppp^{5'}NmpNmpNp$ ($Nm = Am, Gm, Um, Cm$)	m^6A($3m^6A:1m^7G$)
鼠 (BHK-21)	$m^7G^{5'}ppp^{5'}Nmp$ ($Nm = Gm, m^6Am, Um, Gm$)	m^6A, m^5C ($1.8m^6A:1m^7G$)
海　虾	$m^7G^{5'}ppp^{5'}N$	未知
粘 霉 菌	$m^7G^{5'}ppp^{5'}N$ ($N = A, G, Gm$)	无 m^6A
酵　母	$m^7G^{5'}ppp^{5'}Ap$ (75%) $m^7G^{5'}ppp^{5'}Gp$ (25%)	无
丝心蛋白	$m^7G^{5'}ppp^{5'}AmpUmpCp$	m^6A
兔珠蛋白	$m^7G^{5'}ppp^{5'}Nmp$	无 m^6A
鸭珠蛋白	$X^{5'}ppp^{5'}NmpNp$, $X^{5'}ppp^{5'}NmpNmpNp$	无
小鼠免疫球蛋白轻链	$m^7G^{5'}ppp^{5'}N$	
组 蛋 白	$m^7G^{5'}ppp^{5'}N$	

（3）mRNA 分子的编码区——是所有 mRNA 分子的主要结构部分，该区域编码特定蛋白质分子。迄今为止，大量的材料证明除某些已知线粒体外，三联体密码子在所有的生物中是通用的，密码字典是正确的。编码过程和密码子的使用情况详见第六、七两章。

（4）mRNA 中的非编码区——无论原核或真核生物 mRNA 都存在 5′ 端和 3′ 端的两个非编码区。由间隔区转录得到的非编码区广泛存在于原核生物 mRNA 中。在噬菌

体 MS2 和 R17 RNA 的5′末端非编码区129个核苷酸中有三个 AUG 和一个 GUG,但它们都不起起始密码子的作用；在噬菌体 R17 RNA 的外壳蛋白和复制酶基因之间还存在相邻的终止密码子 UAA 和 UAG（见图3-24）。这种现象可能是噬菌体 RNA 为了防止密码子读过头而设置的一种重复终止信号,可以保证蛋白质合成的准确终止。

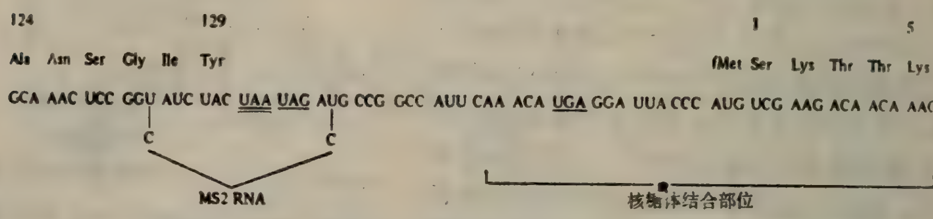


图3-24 R17 RNA 外壳蛋白和复制酶基因之间间隔区的结构

噬菌体 RNA 的非编码区在结构上比其他区域更为保守。QβRNA 的3′端非编码区改变一个核苷酸就会丧失活性。前面已经讲到原核生物 mRNA 编码区的前端及非编码区的一部分是与核糖体结合的部位，图3-25显示 R17 RNA 的复制酶基因起始区和间隔区的一部分形成一个发夹结构,这一部位正是与核糖体的结合点。

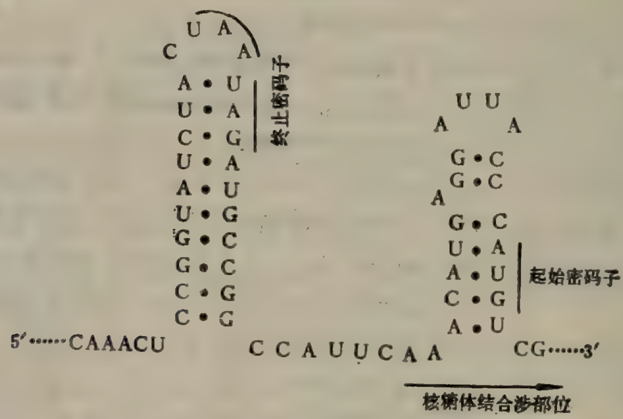


图3-25 R17 RNA 中外壳蛋白和复制酶基因之间间隔区的二级结构

大肠杆菌色氨酸操纵子 mRNA 的 trpB 蛋白和 trpA 蛋白编码区之间的间隔区只有两个核苷酸(UG), trp A 蛋白的起始密码子是 AUG，这样就形成 UGAUG 序列,前三个核苷酸 UGA 起终止密码子的作用,后三个核苷酸 AUG 起起始密码子的作用，这种前后二个密码子共用一个核苷酸的现象是原核生物使用密码的一个特点。某些 mRNA 的间隔区存在着从 DNA 模板上调控区域转录下来的序列，如大肠杆菌色氨酸操纵子 mRNA 的5′端非编码区存在着一段序列是从所谓的衰减子上转录下来的,该“衰减RNA”的3′端是 $CUUUUUUU_{OH}$，这个结构十分相似于在几个大肠杆菌小分子 RNA 和它的噬菌体中发现的序列 $UUUUUUA_{OH}$ 和 $CUUUUU_{OH}$。这些序列被认为是转录终止信号的一部分(见第五章)。

真核生物 mRNA 的3′端非编码区有一段共同序列 AAUAAA，这可能是蛋白质翻译终止的协同信号,也可能与 mRNA 的转录终止有关。在某些真核生物 mRNA 的非编

码区还发现有 m^6A 和 m^5C 等修饰核苷酸,其确切的功能还不清楚。

动植物病毒 RNA

病毒 RNA 包括噬菌体 RNA，植物病毒 RNA，部分动物病毒（RNA 肿瘤病毒）RNA，类病毒 RNA 及由病毒感染寄主后产生的小分子 RNA 等。关于病毒 RNA 的详细知识可参看第八章。噬菌体 RNA 前面也已作了简略介绍。高等动植物病毒 RNA 的结构类似于真核生物的 mRNA，即一般由 5′端帽子结构（表 3-8），5′端非编码区，编码区，3′端非编码区和 3′端聚腺苷酸构成。但有的没有 5′端帽子结构或 3′端聚腺苷酸。一些植物病毒 RNA 有类似 tRNA 结构的 3′端，并能接受氨基酸(详见第八章)。与真核生物 mRNA 一样，高等动植物病毒 mRNA 也是单顺反子结构的，一般讲，一个真核病毒 mRNA 先被翻译成一个大的蛋白质，然后再被裂解成几个较小的具有生物活性的蛋白质。

表 3-8　动植物病毒 RNA 的 5′端和中间甲基化的核苷

RNA	5′端帽子结构	中间甲基化核苷
细胞质多角体病毒 mRNA	$m^7G^{5'}ppp^{5'}AmpGp$	无
呼肠孤病毒 mRNA	$m^7G^{5'}ppp^{5'}GmpCpUp$	无
牛痘病毒 mRNAs	$m^7G^{5'}ppp^{5'}GmpNp$ 和 $m^7G^{5'}ppp^{5'}AmpNp$	无
水泡性口炎病毒 mRNA	$m^7G^{5'}ppp^{5'}AmpAp$ 和 $m^7G^{5'}ppp^{5'}AmpAmp$	无
Sindbis 病毒 “26S” mRNA	$m^7G^{5'}ppp^{5'}Np$ 或 $m^7G^{5'}ppp^{5'}Ap$	m^5C(无 m^6A)
劳氏肉瘤病毒 35S 病毒 RNA	$m^7G^{5'}ppp^{5'}GmpCp$	—
鸟肉瘤病毒 RNA	$m^7G^{5'}ppp^{5'}GmpCp$	m^6A
SV40 mRNA	$m^7G^{5'}ppp^{5'}Amp$ 和 $m^7G^{5'}ppp^{5'}Gmp$	m^6A
腺病毒 mRNA	$m^7G^{5'}ppp^{5'}m^6Amp$	m^6A
Brome mosaic 病毒 RNA4	$m^7G^{5'}ppp^{5'}GpU$	—
黄瓜花叶病毒	$m^7G^{5'}ppp^{5'}Np$	—
烟草花叶病毒	$m^7G^{5'}ppp^{5'}Gp$	—
苜蓿花叶 (Alfalfa mosaic) 病毒 RNA	$m^7G^{5'}ppp^{5'}Gp$	—
人脊髓灰质炎病毒 RNA	一个 22 个氨基酸的多肽	

人脊髓灰质炎病毒 (poliovirus) RNA（参阅第八章）是第一个被测定了全部核苷酸排列顺序的病毒 RNA，由 7,433 个核苷酸构成，具有典型的高等动植物病毒 mRNA 的特点。它具有一个特殊的通过 5′端磷酸根与一个小蛋白共价相连接的 5′端（不同于前面所述的含 m^7G 的帽子结构），随后依次是 5′端非编码区(740 个核苷酸)，编码区 (6,621 个核苷酸)，3′端非编码区(72 个核苷酸) 和 3′端聚腺苷酸。人脊髓灰质炎病毒 RNA 的特殊 5′端见图 3-26，它含一条由 22 个氨基酸构成的肽链，通过酪氨酸上的羟基与其 5′端尿苷酸上的磷酸共价相连接。这个肽段是由病毒 RNA 本身编码的(见表 3-9)。该病毒的

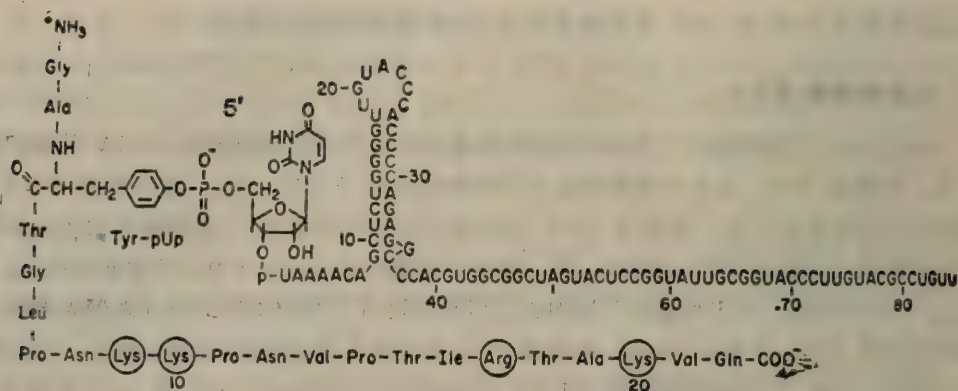


图 3-26 人脊髓灰白质炎病毒 RNA 5′端的结构(为 RNase III 裂解点片段)

5′端非编码区构成一个发夹结构(由 9—36 核苷酸构成),这个区域与调节和辨认的功能有关。而其蛋白质末端可能与 RNA 合成的起始有关。虽然 RNA 上结构基因的翻译起始于大的阅读框架(其起始密码子的位置在 741),但在 741 的前面,即 5′端非编码区,多次出现不作起始信号的 AUG,如位于 185、237、276、318、376、387、455 和 584。这些 AUG 是潜在的起始点,有的后面紧跟着终止信号;其中 185、376 和 584 分别可以作为起始信号而合成分子量为 4,725,8,425 和 7,187 的三个肽段,这些潜在的编码序列的意义还不清楚。

表 3-9 脊髓灰白质炎病毒蛋白质的性质

蛋白质	核苷酸顺序上的位置	氨基酸数	分子量	电荷
病毒肽的编码区	741—7361	2207	246530	+44
Pl-la	744—3380	879	97247	+12
VP0	744—1760	339	37352	+3
VP4	744—947	68	7385	+2
VP2	948—1760	271	29985	+1
VP3	1761—2474	238	26410	0
VP1	2475—3380	302	33521	+9
P2—3b	3381—5105	575	64953	+16
P2—5b	3828—5105	426	48273	+15
P2—X	4119—5105	329	37555	+15
P3—1b	5106—7361	752	84234	+16
VPg*	5367—5432	22	2354	+4
P3—2	5433—7361	643	72132	+11
P3—4b	5979—7361	461	52481	+6
P3—7c	5433—5978	182	19669	+5

* 为病毒 RNA 5′端连接的小蛋白(22 肽)。

编码区是由2207个连续的三联体密码子共6621个核苷酸构成，占整个病毒分子核苷酸数目的89%，被翻译产生一个大蛋白，然后裂解成有生物功能的肽段（见表 3-9）。裂解点通常位于谷氨酰胺和甘氨酸的连接处。其翻译终止信号是由连续的二个终止密码子构成的（UAG 和 UAA）。

3′端非编码区由 72 个核苷酸组成，缺乏在真核生物 mRNA 中存在的 AAUAAA 顺序。这个六核苷酸的缺乏说明该病毒 RNA 的聚腺苷酸是由病毒负链 RNA 的 5′ 端聚尿苷酸转录产生的。该病毒 RNA 的聚腺苷酸长度不是均一的，其平均链长为 60。 聚腺苷酸的功能可能与负链 RNA 合成的起始有关。

脊髓灰质炎病毒 RNA 核苷酸顺序是用测定其 cDNA 顺序的方法推测得到的。最近报道用同样方法也测出了烟草花叶病毒 RNA 的全部核苷酸排列顺序。

类病毒 RNA

类病毒是一类高等植物的病原体，通常由200—400 个核苷酸构成，分子量约 12 万，已经发现的类病毒有十多种。

关于类病毒的一般知识详见第八章，这里就类病毒的结构简介如下：Diener 等人 1967 年发现马铃薯块茎纺锤状病的病原体是一种无蛋白外壳的裸露的 RNA 分子(其沉降系数为 10S)。这种低分子量的 RNA 一旦进入宿主细胞，便在其中复制。Diener 称这种具有感染活性的 RNA 为类病毒。电子显微镜观察马铃薯块茎纺锤状病类病毒(PSTV)，其分子平均长度为 50nm，具有近似 DNA 宽度（20 Å）的棒状结构。在变性条件下，电镜观察到的类病毒呈现网球拍状，有缺口的单链环状分子。多核苷酸激酶在 [γ-^{32}P]-ATP 存在时不能标记天然 PSTV，高碘酸氧化 PSTV 的实验表明它没有游离的 3′ 端邻位羟基。这些结果说明类病毒是单链环状的 RNA 分子。1978 年 Gross 小组报道了由 359 个核苷酸构成的 PSTV 的一级结构(见图 3-27)。这个结果证实了 PSTV 是一个单链环状分子。整个分子具有高度的二级结构，359 个核苷酸中约有 244 个核苷酸处

图 3-27 马铃薯块茎纺锤状病类病毒（PSTV）的结构

在碱基配对之中（占 68%），分子像一根僵直的棒。在整个序列中没有发现真核生物 mRNA 通常使用的起始密码子 AUG，但 GUG 却有七处。终止密码子（UAA、UAG 和 UGA）有六处。类病毒分子如此之小，以至于所带的信息量仅够编码一个蛋白质。它进入寄主以后如何复制自己，如何干扰寄主的正常复制机制，它本身的结构的生物学意义是什么，感染活性的结构基础又是什么？所有这些都有待于进一步探讨。迄今已经测定了一级结构的类病毒还有菊花矮缩病，鳄梨日灼病，椰子 Cadang Cadang 病和柑桔脱皮病等的类病毒，它们都是单链环状的 RNA 分子，其结构特征与 PSTV 相同。

二、RNA 结构研究有关的工具酶

核酸一级结构研究的发展历史说明，无论是 RNA 还是 DNA 的一级结构的阐明都离不开巧妙地使用各种工具酶。六十年代 RNA 一级结构研究所以取得飞跃发展，重要条件之一是许多作用于 RNA 的工具酶的相继提纯和它们的专一性的阐明。一些重要工具酶，例如牛胰核糖核酸酶（RNase A）、核糖核酸酶 T_1（RNase T_1）及蛇毒磷酸二酯酶等的专一性、物理化学性质和化学结构等都已作了详细的研究。七十年代以来，DNA 一级结构研究的突破在很大程度上也取决于大量有序列特异性的限制性内切酶的发现、提纯和实际应用(详见第二章)。因而工具酶的研究是核酸结构研究不可缺少的一个重要方面。下面就已用于 RNA 结构分析的一些比较重要的工具酶作一简要介绍(见表 3-10)，内容主要涉及与结构分析有关的一些问题。

核糖核酸酶（RNase）

RNase 是一类水解 RNA 的酶的通称，它们广泛存在于所有的动、植物和微生物之中。目前已用于结构研究的有 RNase A、RNase T_1、RNase U_2、RNase Phy M 及 RNase T_2 等。根据水解专一性，RNase 可以分为碱基特异性的 RNase 和非特异性的 RNase。上面提及的前四种为碱基特异性的 RNase，后一种为非特异性的。

RNase A　RNase A 是具有碱基特异性的 RNase，它专一性地水解 RNA 链中的嘧啶核苷酸键，产生 3′嘧啶核苷酸和以 3′端为嘧啶核苷酸的寡核苷酸。它水解胞嘧啶核苷酸键的速度大于尿嘧啶核苷酸键的速度。也水解许多碱基上被修饰的嘧啶核苷酸（见表 11）。有人曾提出 RNase A 的作用机制，认为该酶对水解部位的主要要求是：(1) 嘧啶碱的 C_2 为氧等负电性原子；(2) N_3 必须有氢质子。

为了进一步提高 RNase A 的碱基专一性，有人使其肽链第 41 位上的赖氨酸羧甲基化形成羧甲基 RNase A（CM RNase A），羧甲基 RNase A 在适当条件下只水解胞嘧啶和腺嘌呤核苷酸（—CpA—）之间及尿嘧啶和腺嘌呤核苷酸（—UpA—）之间的磷酸二酯键，对—CpG—和—UpG—的水解极慢，而对两个嘧啶核苷酸之间的磷酸二酯键则不能水解。

RNase A 的水解专一性不是绝对的，当它在高酶浓度时也水解聚腺苷酸，生成 3′腺苷酸。

RNase T_1　此酶是从一种米麹(霉菌)中制得的，具有高度的碱基专一性，能特异性地裂解 RNA 产生 3′鸟苷酸和 3′端为鸟苷酸的寡核苷酸。该酶也能水解肌苷酸(Ip)和部

表 3-10 常用工具酶及其性质*

名 称	类 型	底 物	产 物
RNase T_1	内 切	—GpN—	Gp、—Gp
RNase A	内 切	—PypN—	Pyp、—Pyp
RNase U_2	内 切	—PupN—	Pup、—Pup
RNase T_2	内 切	—NpN—	Np
RNase Phy I	内 切	—ApN—,—GpN— —UpN—,—CpN—	Ap,—Ap,Gp,—Gp Up, —Up, 极慢产生 C>P 和—C>P
RNase Phy M	内 切	同 上 (限制性条件下只水解 —ApN—, —UpN—)	同 上 (Up,—Up,Ap,—Ap)
RNase M	内 切	—NpN—	Np
核酸酶 P_1	外切、内切	DNA、RNA	pN
核酸酶 M	内 切	DNA、RNA	Np
核酸酶 S_1	内 切	单链 DNA 或 RNA	pN
蛇毒磷酸二酯酶	外 切	RNA、DNA	pN
牛脾磷酸二酯酶	外 切	RNA、DNA	Np
PNPase	外 切	Pi+—NpN—	ppN**
磷酸单酯酶		pNpNp	2Pi+NpN
多核苷酸激酶		ATP+ $_{HO}$NpN—	pNpN—+ADP

* Py、Pu 分别代表嘧啶核苷和嘌呤核苷,N表示任一种核苷。

** PNPase 为多核苷酸磷化酶,它磷酸解 RNA 得到的最终产物中还伴有二核苷一磷酸或三核苷二磷酸。

分碱基上被修饰的鸟苷酸键(见表 3-11),其中水解 m_2^2Gp 键的产物是 2′、3′-环化的m_2^2Gp ($m_2^2G>P$)。

有人提出 RNase T_1 对底物的要求是嘌呤环 6 位上的酮基和 7 位上的三价氮原子。当 N-1 存在有一个质子时,将更易受酶的攻击。因此这些位点被认为是酶的专一性结合点。这个假说可以解释能被 RNase T_1 水解的许多修饰核苷酸,但 1-甲基肌苷酸 (m^1Ip) 键不能被酶解是个例外。

表 3-11 RNase A 和 RNase T_1 的专一性

名 称	水解 RNA 后产生的核苷酸	不能水解的核苷酸键
RNase A	Cp、Up、m^5Cp、ac^4Cp、s^2Cp、m^5Up(Tp)、cmo^5Up、mnm^5s^2Up、m^5S^2Up、s^4Up、Dp、ψp	m^3CpN、m^3UpN、ncp^3UpN
RNase T_1	Gp、Ip、Xp、m^1Gp、m^2Gp、m_2^2Gp	m^7GpN、m^1IpN、QpN、WpN

* 符号和结构分别见第一章和附录一。

RNase U_2　　该酶是从一种黑粉菌 (*Ustilago sphaerogena*) 的培养液中提取的一种对嘌呤核苷酸键专一的 RNase。其水解专一性正好与 RNase A 互补,但产物往往是 2′, 3′

环状的嘌呤核苷酸，与 RNase T_1 相比较，RNase U_2 的水解专一性不是绝对的，其水解二核苷一磷酸的速度依次是 A—N > G—N ≫ C—N > U—N。在结构分析中使用该酶时常常先将 RNA 分子上的鸟嘌呤碱基用特异性的化学试剂(例如水溶性碳二亚胺等)化学修饰，修饰后的鸟苷酸键不受 RNase U_2 的攻击。这样 RNase U_2 只水解 RNA 链上的腺苷酸键。除 A 和 G 外，该酶还能水解 I 和 V，但不能水解 X、i^6A，ms^2i^6A，t^6A、m^7G 和 m_2^2G 等修饰核苷酸键。

RNase T_2　与 RNase T_1 一样，RNase T_2 也是从米麴中制得的一种 RNase，不过这是一种非特异性的 RNase。它水解 RNA 的最适 pH 在 4.5，水解各种 2′，3′环状核苷酸的最适 pH 在 6.0—6.3；水解腺苷酸键的速度最快，2′，3′环状腺苷酸几乎不积累，显示腺苷酸特异性内切酶的性质。水解各核苷酸键的速度依次是 A ≫ U > G > C。它能水解所有的修饰核苷酸键，包括 RNase T_1 和 U_2 不能水解的 m^7G 等，但不能水解 CMC-G 和 CMC-U（CMC-为水溶性碳二亚胺修饰基团）以及所有的 2′-0-甲基化的核苷酸键。底物分子的构型对 RNase T_2 的作用是敏感的，在水溶液中的聚腺苷酸与 RNA 分子上的腺苷酸比较是抗 RNase T_2 的，但在 40% 甲醇中对聚腺苷酸的水解速度明显加快，可见具有高级结构的 RNA 分子是抗 RNase T_2 的。

RNase Phy I [多头绒孢菌核糖核酸酶 I]　这是一个特异性不强的 RNase，它能很快地水解聚尿苷酸和聚腺苷酸，产生 2′，3′环状单核苷酸，然后很慢地水解环状单核苷酸成为 3′核苷酸；它水解聚胞苷酸的速度极慢，且产物为 2′，3′环状胞苷酸。除 CpN 以外所有的二核苷酸均能很快地被该酶水解。由于它水解胞苷酸键的速度特别慢，人们已将它在结构分析中与 RNase A 相配合使用以区分两个嘧啶核苷酸 C 和 U（见第 101 页）。

RNase phy M（多头绒孢菌核糖核酸酶 M）　该酶的性质与 RNase phy I 在通常酶解条件下完全一样，即其水解二核苷一磷酸的速度次序如下：UpN > ApN > GpN > CpN；NpA > NpC > NpG > NpU。因此 UpC、UpA、ApA 和 ApC 最易水解，而 CpU、CpC 最不易水解。但在限制性条件下（7M 尿素，50℃，pH 5.0）该酶只水解 UpN 和 ApN。因而较之 RNase Phy I 更有利于 RNA 序列分析(图 3-33)。

RNase M　该酶是从一种麴霉菌（*Aspergillus saito*）中提取的非特异性 RNase，水解 RNA 时释放各核苷酸的速度的次序是 A ≫ U > C > G，最适 pH 是 4.0，其活性能被二价铜离子和二价锌离子所抑制。

其他 RNase　用于 RNA 结构分析的 RNase 还有 RNase U_1，BS RNase（枯草杆菌 RNase）和 RNase N_1（红色链孢霉 RNase）等。RNase U_1 的性质与 RNase T_1 极相似，是鸟嘌呤特异性的 RNase，其最适 pH 在 8.0—8.5，可代替 RNase T_1 用于 RNA 的结构分析。RNase N_1 的水解专一性与 T_1 和 U_1 相似，但它的转磷酸活性比水解活性大得多，水解产物往往是鸟嘌呤 2′，3′环状核苷酸，延长水解时间也可以产生 3′鸟苷酸。BS RNase 是一种非特异性的 RNase，它水解 RNA 的产物往往是 2′，3′环状核苷酸。

核酸酶（Nuclease）

这是一类既能水解 RNA 也能水解 DNA 的酶，其水解特性类似于磷酸二酯酶，产物通常为单核苷酸。这类酶已被用于结构分析的有核酸酶 P_1、核酸酶 M 和核酸酶 S_1。

核酸酶 P_1　核酸酶 P_1 是从桔青霉（*Penicillium citrinum*）中提取并提纯的一个工

具酶。它水解核酸产生 5′核苷酸；没有碱基特异性；最适 pH 4.5—6.0，最适温度 70℃；降解核酸时需要锌离子。这个酶兼有 3′核苷酸酶的活性，即它能水解核苷酸或寡核苷酸的 3′ 端磷酸；作为 3′核苷酸酶时它水解核糖核苷 3′磷酸的速度比水解脱氧核糖核苷 3′磷酸的速度快 30 倍。由于它具有良好的水解核酸的活性，来源方便，已被广泛地用于核酸研究的各个方面。

核酸酶 M　　核酸酶M是从金黄色葡萄球菌培养液中制得的一种核酸酶。它能水解核酸产生 3′核苷酸；它水解变性 DNA 的速度快于天然 DNA；能水解 RNA 但不能水解核糖核苷 2′，3′环化磷酸酯，这说明它的水解机制不经过 2′，3′环状磷酸的中间阶段；该酶被带有 5′磷酸根的单核苷酸或寡核苷酸抑制； 3′磷酸根的存在却能促进邻近核苷酸键的水解；它水解产生 Ap 和 Up 的速度大于产生 Cp 和 Gp 的速度；如果用二价锶离子代替二价钙离子作激活剂，则它只能水解 DNA 而不能水解 RNA。由于它水解 RNA 的产物与 RNase T_2 的相似，因此可替代 RNase T_2 作组成分析。

核酸酶 S_1　　核酸酶 S_1 是从一种麴霉菌 (*Aspergillus oryzae*)提取的一种单链核酸酶。它只水解单链核酸，不能水解天然的双链 DNA； 产物是 5′核苷酸；它水解 DNA 和 RNA 有相同的最适 pH (pH4.5)； 它的水解 DNA 的活性能被 RNA 所抑制。该酶已被广泛地用于单链核酸的检定和核酸高级结构的研究等方面。

外切核酸酶 (exonuclease)

外切核酸酶是指一类水解核酸(包括 DNA 和 RNA) 时是从核酸分子的一端开始逐个地把核苷酸顺次水解下来的核酸水解酶。用于结构分析的这类酶有蛇毒磷酸二酯酶，牛脾磷酸二酯酶和多核苷酸磷酸化酶 (PNPase)。

蛇毒磷酸二酯酶　　它水解核酸是从 3′端开始将 5′单核苷酸逐个地顺次水解下来；既能水解 DNA 也能水解 RNA，产物都是 5′核苷酸；它水解变性 DNA，也水解天然双链 DNA，而且速度更快；对单链 RNA 的水解能很快完成，对于水解双链 RNA 虽有不同的说法，但一般认为高级结构对蛇毒磷酸二酯酶的水解活性没有明显的影响；它能水解 2′-O-甲基化的核糖核苷酸键，但水解速度比没有甲基化的慢一些，它还能水解一个或二个核苷都是阿拉伯糖苷的二核苷一磷酸，可见糖基的性质对蛇毒磷酸二酯酶的水解活性影响不大(这个性质与核酸酶M不同，核酸酶M不能水解带有阿拉伯糖基的核苷酸键)；它对常见的四种核苷酸没有明显的碱基专一性，但对某些稀有核苷酸的水解速度与常见核苷酸的不同，它水解 m_2^2GpCp 比 GpCp 慢得多，水解 Tpϕ 非常慢，水解 Cpϕ 的速度只有 CpU 的一半。经水溶性碳二亚胺修饰后的尿苷酸键则不能被蛇毒磷酸二酯酶水解，可见它水解核酸的活性也有一定的碱基专一性；它对被水解的核酸的末端磷酸基有严格的要求，同一核苷酸片段如果 5′端带磷酸根的则水解最快，去除 5′端磷酸根的则水解速度降低 10 倍；如果 3′端带有磷酸根的则水解速度降低 100 倍(比 5′端带有磷酸根 3′ 端不带磷酸根的则慢 1,000 倍)。可见 5′端的磷酸根促进水解，3′端磷酸根抑制水解。

牛脾磷酸二酯酶　　此酶也是一个外切核酸酶，它水解 RNA 或 DNA 是从核酸分子的 5′端开始逐个顺次地把 3′核苷酸水解下来；在 pH 4.8 时它不能水解 5′端带有磷酸根的 tRNA，但它能很好地水解 5′端脱去磷酸根的 tRNA；在 pH 6.2 时，它能够水解 5′端带有磷酸根的 tRNA， 但其水解速度低于脱磷的 tRNA，可见 5′端磷酸根的存在会抑

制或降低其水解速度；它的水解活性对底物的二级结构敏感，它水解天然胸腺 DNA 的速度只有变性 DNA 的百分之四，双链 DNA 片段（20—50 个碱基对）不能为该酶作用，在水溶液中能形成双链结构的，如聚肌胞苷酸等完全不能为该酶水解；它能水解 RNA 分子产生 2′-O-甲基化的核苷酸和许多碱基上修饰的核苷酸，但不能水解 2′-O-乙酰化的核苷酸键，也不能水解核苷 2′，3′环状磷酸。

多核苷酸磷酸化酶　PNPase 是一个具有外切核酸酶性质的核苷酸转移酶。这个酶实际上催化三个反应：(1)以 5′核苷二磷酸为底物使其聚合成多核苷酸，同时释放出无机磷酸；(2)以多核苷酸为底物，在无机磷酸存在下使多核苷酸磷酸解产生 5′核苷二磷酸（为 1 的逆反应）；(3) 催化无机磷酸交换反应。它在催化磷酸解多核苷酸时是从 3′端开始逐个顺序地把核苷酸水解下来，产物为 5′核苷二磷酸；3′端磷酸根抑制磷酸解反应；它磷酸解多核苷酸链最后得到该多核苷酸 5′端的二核苷一磷酸或三核苷二磷酸；它对底物无明显的碱基专一性，既作用于 RNA 也作用于 DNA。PNPase 曾被用于 RNA 或 DNA 片段的人工合成（见第十章）。

其他有关的酶

核酸结构分析中常用的工具酶还有磷酸单酯酶（PMase），多核苷酸激酶（PNKase）和 RNA 连接酶等。

磷酸单酯酶　PMase 有非特异性磷酸单酯酶和特异性磷酸单酯酶二类。非特异性磷酸单酯酶如大肠杆菌碱性磷酸酯酶，它能够水解单核苷酸或寡核苷酸的 3′ 或 5′ 端磷酸单酯键，产生无机磷。它也能水解核苷-2′一磷酸单酯、核苷二磷酸、核苷三磷酸、糖磷酯、对硝基酚磷酸和焦磷酸等范围广泛的磷酸单酯化合物。此外，它还能水解硫磷酯。大肠杆菌碱性磷酸单酯酶有很高的催化活性，已被广泛地用于核酸研究的许多方面；它已被制成固相酶，使用和除去更为方便。

非特异性磷酸单酯酶还有哺乳动物的肠粘膜、肾脏、肝脏和胎盘的碱性磷酸酯酶及许多其他来源的酸性磷酸酯酶，例如前列腺、红血球、肝脏、脾脏、血清等来源。

特异性磷酸单酯酶与核酸结构分析有关的有 5′核苷酸酶和 3′核苷酸酶，它们分别特异性地水解核苷酸或寡核苷酸的 5′磷酸单酯键和 3′磷酸单酯键，产生无机磷酸。5′核苷酸酶的来源有脊椎动物组织、精液、蛇毒、猪血和微生物等许多方面，其中已用于核酸结构分析的有大肠杆菌 5′核苷酸酶，蛇毒 5′核苷酸酶等。它们的共同特性是既水解核糖核苷 5′磷酸也水解脱氧核糖核苷的 5′磷酸。大肠杆菌 5′核苷酸酶还能水解 ATP、UTP、CTP、GTP 和 UDP-葡萄糖等，它水解核糖核苷单磷酸的相对速度是：A ＞ G ＞ C ＞ U ＞ I。蛇毒 5′核苷酸酶水解核糖单核苷酸的相对速度是：A ＞ C ＞ I ＞ G ＞ U。

3′核苷酸酶有黑麦草、麦苗、绿豆和微生物等多种来源的制剂。它们特异性地水解核苷酸或寡核苷酸的 3′磷酸。

多核苷酸激酶　T4 噬菌体感染大肠杆菌产生的多核苷酸激酶能催化 ATP 上的 γ 磷酸根转移到核酸或寡核苷酸片段以及 3′核苷酸的 5′羟基上。它能够作用于 DNA 亦能作用于 RNA；作用于 DNA 的速度比 RNA 快；它催化各种寡核苷酸磷酸化的速度取决于寡核苷酸的链长、组成及其糖基的种类。一般情况下，它磷酸化 3′单核苷酸的速度是寡核苷酸的 2 到 5 倍。T4 噬菌体感染的大肠杆菌多核苷酸激酶伴有 3′核苷酸酶的活性，

它的激酶活性的最适 pH 在 6.5—8.5，而 3′核苷酸酶的最适 pH 在 5.9，该酶水解脱氧核糖核酸 3′ 磷酸的速度比水解核糖核酸的更快些；已经有人从 T4 噬菌体突变株感染的大肠杆菌中制得了不含 3′核苷酸酶活性的激酶。多核苷酸激酶已被广泛用于核酸的结构研究。

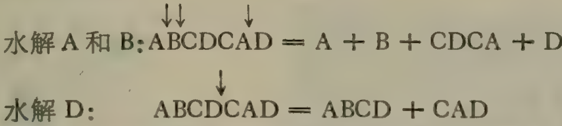
RNA 连接酶 （详见第十章）。

三、RNA 一级结构测定方法

RNA 一级结构测定方法可以分为两大类：片段重叠法和直接阅读法。前者为经典方法，现在仍然在使用；后者是由 DNA 顺序分析衍生而来的快速简便的新技术，它问世仅几年时间，已经使核酸一级结构的研究出现了崭新的局面。

片段重叠法

重叠法是测定包括蛋白质在内的生物大分子一级结构的一种战略，其原理如下：RNA 分子分别被二种以上的具有不同碱基专一性的核糖核酸酶水解成各种寡核苷酸片段；然后分别分离和检定水解产物，并测定它们的核苷酸顺序及其相对含量；二种不同专一性的酶得到的二组寡核苷酸的切点是相互重叠的。利用这个原理，使用更多种专一性不同的酶水解 RNA 分子，解析结构的能力就更强。这就是片段重叠法测定一结构的基本原理。例如我们现在要解析 ABCDCAD 这个假定的顺序，有两种酶可供使用，一种专水解 A 和 B；另一种专水解 D。结果如下：



水解 A 和 B：ABCDCAD ＝ A ＋ B ＋ CDCA ＋ D

水解 D：　ABCDCAD ＝ ABCD ＋ CAD

这样就可以推断这个片段的序列是 ABCDCAD。因为第一个酶解产物中有 CDCA 序列，说明该序列在片段中是存在的；第二个酶水解得到 ABCD 和 CAD 二个片段，其排列有二种可能性 ABCDCAD 和 CADABCD，因为其中有 CDCA 顺序，所以应该是前一种序列。

重叠法的基本实验方法和技术。

碱基组成分析　碱水解法：RNA 经碱水解产生 2′ 和 3′ 核苷酸。注意下列核苷对碱不稳定：D、m^1A、m^3C、m^7G、m^1I、ac^4C、V 等。酶解法：利用非特异性核糖核酸酶（RNase T_2 或 M）水解，产物是 3′核苷酸。上述水解产物可经电泳或双向电泳层析进行组成分析（见附录二）。

末端分析　5′端和 3′端核苷酸目前一般用同位素标记的方法进行测定。5′ 端同位素 ^{32}P 标记用多核苷酸激酶在 [γ-^{32}P]-ATP 存在下使 RNA 片段 5′端羟基磷酸化来实现。3′端也可以用 ^{32}P 进行同位素标记，通过 RNA 连接酶催化 [^{32}P]-pCp 与被测寡核苷酸之间的单加反应来实现。末端同位素标记的寡核苷酸经核酸酶降解后，用测定核苷酸组成的方法确定带有同位素[^{32}P]的 3′或 5′端。

特异性酶解和酶解产物的分离　通常使用的二种酶是 RNase T1 和 RNase A，前者对鸟苷酸键专一，后者对嘧啶核苷酸键专一。得到的全酶解产物可通过 7*M* 尿素柱层

析进行分离（见图 3-28）。中性尿素柱层析可以按链长把十核苷酸以下的寡核苷酸片段分开。该法一般用于非放射性样品的分离。

图 3-28 DEAE-纤维素柱层析分离寡聚核苷酸混合物。洗脱溶剂为 pH 6.5 碳酸氢三乙胺梯度，内含 7*M* 尿素。图中数目字表示链长。

此外，还可以用双向电泳法分离同位素标记的酶解产物，有人把这种方法称为指纹法。图 3-29 显示了用 Sanger 建立的双向电泳法分离放射性[^{32}P]标记的核酸水解产物的图谱，该法简便易行，已广泛地用于 RNA 一级结构特别是小分子 RNA 的测定。较大片段的分离一般采用聚丙烯酰胺凝胶电泳进行。

图 3-29 用双向电泳法分离 RNA 酶解产物图（蓖麻蚕后部丝腺体 $tRNA^{Gly}$ 的 RNase T_1 全酶解）
第一向：醋酸纤维素薄膜电泳，pH 3.5（醋酸-吡啶）；
第二向：DEAE 纸电泳，7%甲酸。

图 3-30 打点法测定寡核苷酸片段的核苷酸排列顺序
第一向：pH 3.5 醋酸纤维素电泳；
第二向：同系层析。

小片段寡核苷酸的顺序测定

(1) 蛇毒磷酸二酯酶部分酶解产物的打点法——片段的 5′端经 [^{32}P] 标记后，蛇毒磷酸二酯酶部分酶解，双向电泳层析分析(第一向电泳按碱基组成分离；第二向同系层析，按链长分离)，根据同位素斑点在图谱上顺次排列的轨迹即可推导出核苷酸的排列顺序（图 3-30)，本法也常用于小片段 DNA 顺序的测定(见第二章第 17 页)。

(2) 高碘酸氧化法——先用高碘酸氧化 RNA 3′ 端邻位羟基，然后在弱碱性条件下发生 β 消除反应，测定 β 消除反应产生的碱基，即可知道 3′端第一位核苷酸。切除第一个核苷酸的寡核苷酸经磷酸单酯酶处理切去 3′磷酸根后就可进入第二循环(见图 3-31)。由此逐个测定其产生的碱基即可排出该寡核苷酸的序列。目前文献报道最多的循环次数达 29 次。该法不能用于带有 2′-O-甲基化的寡核苷酸片段；高碘酸对假尿苷和二氢尿苷有破坏作用，使用时要注意。产物除用紫外吸收法测定碱基外，还可以用刘望夷等人提出的

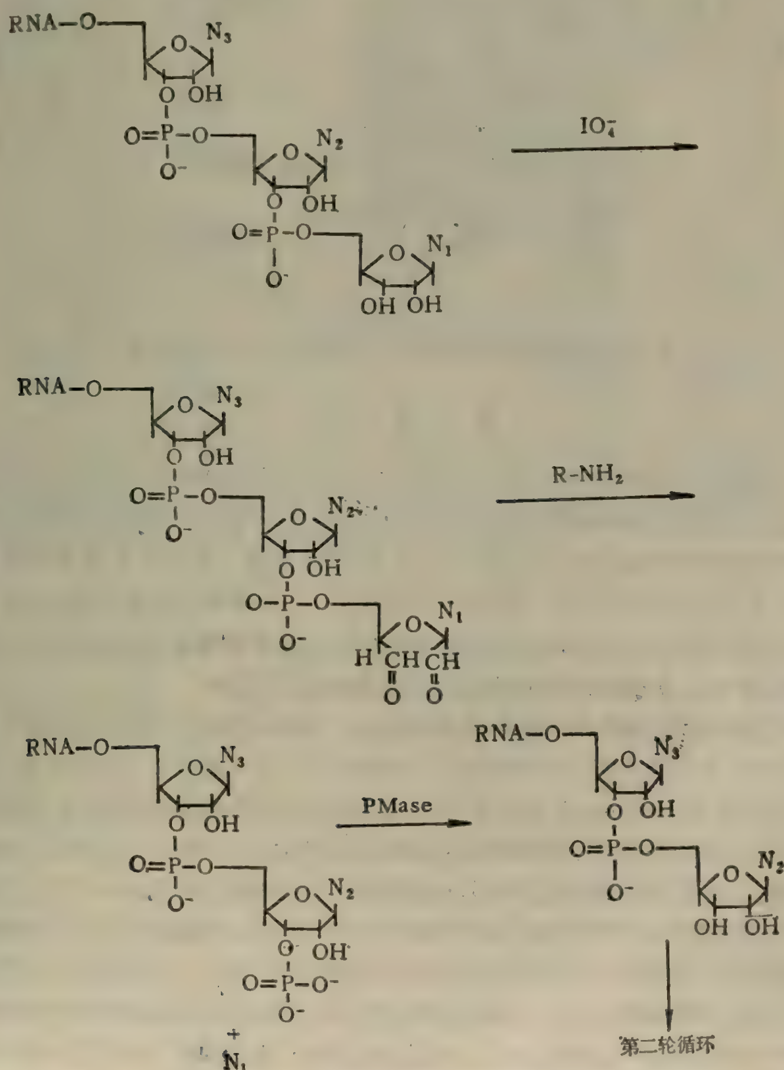


图 3-31 高碘酸氧化法测定寡核苷酸的顺序

萤光法进行检测，此法的原理是高碘酸氧化后产生的二醛与带有萤光基团的酰肼反应，产生萤光性的核苷腙或次甲基核苷腙(图 3-32)，检测核苷腙的萤光性比上述紫外方法灵敏得多。

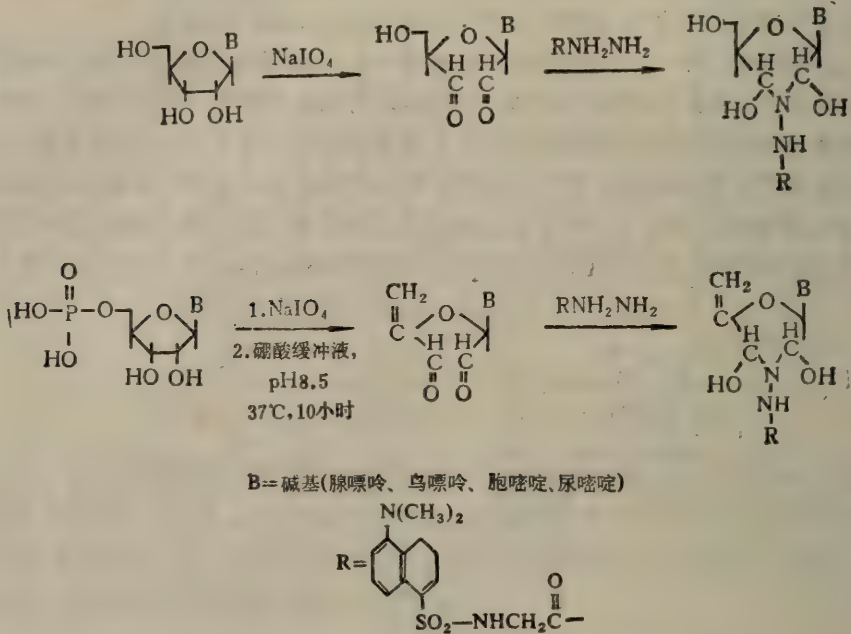


图 3-32 用萤光法检测 DNS-Gly-核苷腙或一次甲基核苷腙

直 读 法

关于直读法的原理详见 DNA 结构一章。利用特异性降解或在控制条件下合成得到的以 RNA 分子中某种特定碱基(如 A) 为末端(3′端或 5′端)的不同链长的寡核苷酸片段混合物，经变性聚丙烯酰胺凝胶电泳按链长分离，凝胶电泳上显示的各条带的链长就表示 RNA 分子上各个 A 碱基的位置。同样，亦可得 G、C 和 U 接尾的各组片段，并定出它们在 RNA 分子中的位置。由此可以很快排出 RNA 分子的核苷酸顺序(见图 3-33)。该方法的特点是简便快速。根据阅读的方法不同又可分为以下几种:

利用专一性核糖核酸酶控制酶解的直读法　某些 RNA 水解酶具有特定的碱基专一性，如前介绍: RNase A 专一水解 C 和 U; RNase T_1 专一水解 G; RNase U_2 在适当条件下专一水解 A; RNase phy I 能水解 A、U 和 G，水解 C 的速度特别慢; RNase phy M 在适当条件下专一水解 A 和 U; 葡萄球菌核酸酶在 pH7.5，10mM Ca^{++}条件下，水解 N-A 和 N-U 很快; 小球菌核酸内切酶在 7M 尿素存在时能水解 C—N 键。如果 RNA 分子的 5′端被^{32}P 标记，适当选择几种酶相配合进行控制性水解。通过聚丙烯酰胺凝胶电泳的分离，可得到几组以 5′端为^{32}P 而 3′端为某一特定碱基的片段。放射自显影后即可阅读顺序(见图 3-33)。此方法受 RNA 二级结构影响较大，同时也不能阅读修饰核苷酸。

利用专一性化学降解的直读法　RNA 在 RNA 连接酶的作用下，在其 3′端接上^{32}pCp 后分别用硫酸二甲酯(修饰 G)、焦碳酸二乙酯 (主要修饰 A)、肼(修饰 U)、无

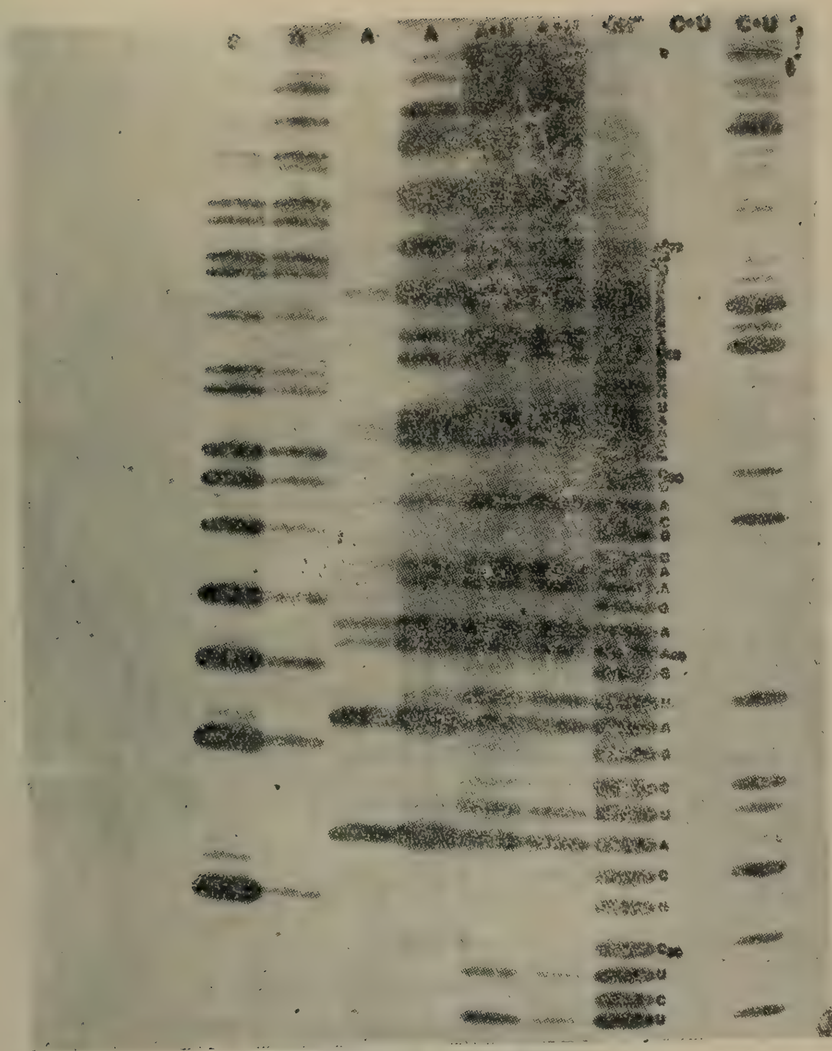


图 3-33 酶解阅读法测定 RNA 结构凝胶电泳图

水肼(在 3*M* NaCl 中主要修饰 C) 进行特异性碱基的限制修饰(其原理可参见 DNA 结构一章),得到几组以 3′端为^{32}P 标记,5′端表示特定碱基的片段,用变性凝胶电泳分离各片段并阅读顺序。该方法的优点是不受 RNA 二级结构的影响,但也不能阅读修饰核苷酸。

利用 RNA 复制特异性末端终止的直读法 在有四种核苷三磷酸存在时,以寡聚核苷酸作引物,依赖于 RNA 的 RNA 聚合酶能在 RNA 模板上合成互补的 RNA 链。如果在反应体系中加入某种一定量的 [α-^{32}P]-3′-脱氧核苷三磷酸,则 RNA 合成就会在这一核苷酸位置上终止,结果产生一系列以引物为 5′端,以 [^{32}P] 标记的 3′脱氧核苷酸为 3′端的不同链长的 RNA 片段。利用 A、G、C、U 四组不同 [α-^{32}P]-3′-脱氧核苷三磷酸的反应体系,经变性凝胶电泳分离,即可阅读其核苷酸顺序。这种方法受到[α-^{32}P]-3′-脱氧核苷三磷酸和依赖于 RNA 的 RNA 聚合酶的来源的限制。另外,如果 RNA 链上有稀有核苷酸,那么 RNA 复制就会受到影响而不能阅读出 RNA 的全序列。这种直读



图 3-34 利用末端鉴定直读法测定 RNA 顺序示意图

○,同位素标记; ●,放射自显影斑点。

法目前应用较少。

利用末端鉴定的直读法 RNA 在无离子甲酰胺或双蒸水的限制性水解条件下,从统计角度而言,每一分子 RNA 只有一个磷酸二酯键被水解。因水解无碱基专一性,得到一组 3′端的片段和一组 5′端的片段。如果原来的 RNA 分子 5′端带有磷酸基团,那么只有 3′端的一组片段的 5′端有游离羟基,这组片段经 5′端标记^{32}P 后在聚丙烯酰胺凝胶电泳上按链长分离。切下凝胶条带或转移至 DEAE-纤维素薄板上,经水解鉴定它们的 5′端的核苷酸,就可排出 RNA 的核苷酸序列(图 3-34)。该方法的优点是能克服 RNA 分子二级结构对测定的影响也能测定稀有核苷酸,缺点是对多核苷酸激酶的质量要求很高,且操作比较麻烦。

反转录成 cDNA 的直读法 在有四种脱氧核苷三磷酸存在时,用反转录酶或大肠杆菌 DNA 聚合酶在 RNA 样板上反转录成 cDNA。然后用 DNA 直读法测定 cDNA 的序列再推导出 RNA 的序列。DNA 序列测定法详见 DNA 结构一章。目前许多 mRNA 和病毒 RNA 都是用这种方法测定序列的。

近年来许多核酸分子一级结构的测定大都是用快速直接阅读法完成的，然而经典的片段重叠法仍不失其特点，它的一套基本实验方法，包括末端分析、组成分析、修饰核苷酸分析等，仍然是 RNA 特别是 tRNA 以及 snRNA 顺序测定的常用方法。

参 考 文 献

[1] Schimmel, P. R., Söll, D. and Abelson, J. N. (eds.), Transfer RNA: Structure, Properties, and Recognition, 1979, Cold Spring Habor Laboratory.

[2] Kim, S. H., 1978, Three-Dimensional Structure of tRNA and its Functional Implications, in *Advances in Enzymology*, **46**, 279, eds. Meister, A., An Interscience Publication, John Wiley and Sons.

[3] Wittmann, H. G., 1982, Components of Bacterial Ribosomes, in *Ann. Rev. Biochem.* **51**, 155—183.

[4] Erdmann, V. A. et al (eds.), 1983, Collection of Published 5S and 5.8S Ribosomal RNA Sequences, *Nucleic Acids Res.* **11**, r105—r133.

[5] Noller, H. F. and Woese, C. R., 1981, Secondary Stucture of 16S rRNA, *Science*, **212**, 405—410.

[6] Noller, H. F. et al, 1981, Secondary Structure Model for 23S rRNA, *Nucleic Acids Res.* **9**, 6167.

[7] Veldmann, G. M. et al., 1981, The Primary and Secondary Structure of Yeast 26S rRNA, *Nucleic Acids Res.* **9**, 6935.

[8] Fiers, W. et al., 1976, Complete Nucleotide Sequence of Bacteriophage MS2 RNA, *Nature*, **260**, 500—507.

[9] Cohn, W. E. and Volkin, E. (eds.) 1976, mRNA: The Relation of Structure atod Functions, in *Prog. Nucleic Acid Res. Mol. Biol.* Vol. **19** Academic Press.

[10] Kitamura, N. et al., 1981, Primary Structure, Gene Organization and Polypeptide Expression of Poliovirus RNA, *Nature*, **291**, 547—553.

[11] Gross, H. J. et al, 1978, Nucleotide Sequence and Secondary Structure of Potato Spindle Tuber Viroid, *Nature*, **273**, 203—208.

第四章 DNA 复 制

李其樑 徐有成

在细胞分裂过程中，亲代细胞所含的遗传信息源源本本地传递到二个子代细胞。我们知道，DNA 携带了细胞的一切遗传信息，因此亲代 DNA 也必须准确地复制成二个拷贝，分配到二个子代细胞中去。研究 DNA 复制的目的就是要了解(1)子代 DNA 为什么能够真实地获得亲代 DNA 的遗传信息；(2)复制是怎样进行的？(3)生物体是怎样对 DNA 复制进行调控的？

对于前两个问题，Watson 和 Crick 在发表 DNA 双螺旋模型时就进行了探讨。在 DNA 双螺旋模型中，G 只能和 C 配对，A 只能和 T 配对，因此按碱基配对规则，任意一条链的碱基顺序可精确地决定它的互补链的顺序。在 DNA 复制时，只要简单地把 DNA 双链拆开，每一条链都可作为模板，合成结构上完全一样的二条 DNA 双链(图 4-1)。

然而，这样一个表面上看来很简单的过程，并不象最初提出时那样，可以象拉"拉链"

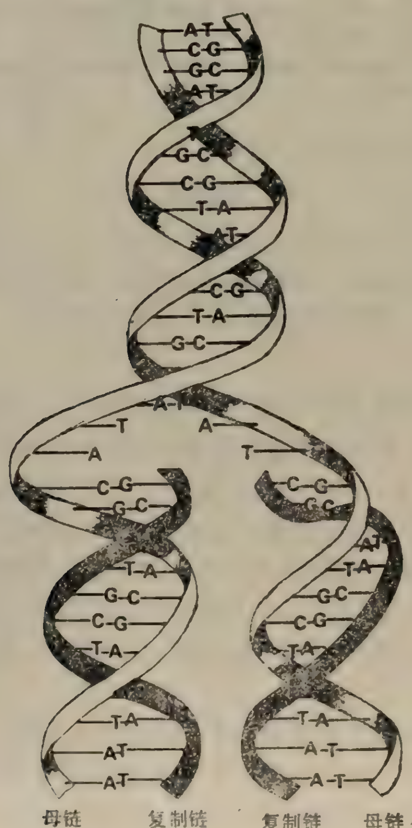


图 4-1 双螺旋 DNA 复制过程

一样不靠任何酶的帮助而自动完成。现在知道，复制过程十分复杂，涉及 30 多种酶和蛋白质的协同作用。表 4-1 列出部分与大肠杆菌复制有关的蛋白质及其性质。其中大部分酶和蛋白质的结构和作用已经弄清楚，对双链如何解开以及怎样保持复制真实性也有了一定的了解，但仍是知其然而不知其所以然。由于复制是生命活动中极为重要的一环，感兴趣的实验室很多，在这个领域里进展很快，因此我们只能把复制的基本过程作一介绍。

表 4-1　与大肠杆菌复制有关的蛋白质

蛋白质名称*	分子量×10^3	亚基数目	功　能	每个菌中的分子数	毫克蛋白质/公斤细胞(湿重)
SSB	74	4	与单链结合	300	20
i 蛋白	80	4	预引发	50	0.5
n 蛋白	25	1	预引发	30	
n′ 蛋白	75	1	部位识别，ATPase	70	0.3
n″ 蛋白	11	1	预引发	—	
dnaC	29	1	预引发	100	
dnaB	300	6	移动启动子，ATPase	20	0.3
引发酶	60	1	引发合成	50	0.2
Pol III 全酶			链的延长		0.5
α	140	1		—	
ε	25	1		—	
θ	10	2		—	
β	40	1		300	
γ	52	2		20	
δ	32	1			
τ	83	1			
Pol I	109	1	填补缺口和切去引物	300	10
连接酶	74	1	连接	300	10
拓扑异构酶 II	400	4	引进超螺旋	—	
α	210	2		250	
β	190	2		25	
rep 蛋白	65	1	解开双链	50	0.6
解链酶 II	75	1		5000	
dnaA	48		复制起始	200	

* SSB，单链结合蛋白；Pol，DNA 聚合酶；rep 蛋白是一种大肠杆菌解链酶。

一、复 制 机 制

半 保 留 复 制

Watson 和 Crick 提出的 DNA 复制方式可称为半保留的复制方式，因为每个子代 DNA 分子中一条链是完整的亲代 DNA 链，另一条链是新合成的链。在理论上还可以设想其他一些复制方式。例如二个子代 DNA 分子中一个分子完全是亲代的分子，而另一个分子完全是新合成的分子，这种方式可以称为全保留的复制方式；又比如可以设想亲代链被分解成很多片段，散布在新的子代分子中，这种方式称为散布式复制机制(图 4-2)。

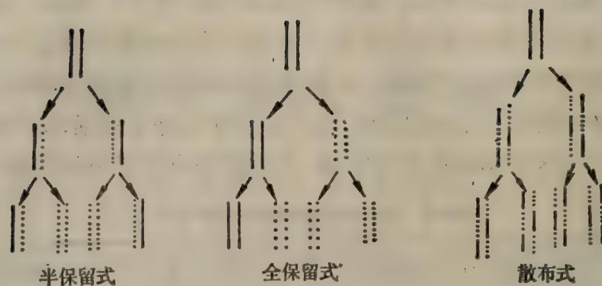


图 4-2 复制机制的三种可能方式

生物体内究竟以什么方式进行复制，必须用实验来证明。1958 年 Meselson 和 Stahl 设计了一个很巧妙的实验。他们先把大肠杆菌在 96.5% 同位素纯度的 $^{15}NH_4Cl$ 培养液中生长十五代，使所有的 DNA 都被重的 ^{15}N 标记。然后把细胞转到含 $^{14}NH_4Cl$ 的介质中培养，在不同的时间取出样品，每个样品都用 SDS 裂解细胞。裂解液放在 CsCl 溶液中以 140,000g 离心二十小时。离心结束，从管底到管口 CsCl 密度形成一个梯度分布，越向上密度越低。DNA 分子的密度如果同某一层 CsCl 密度相等时，它就不上也不下，在

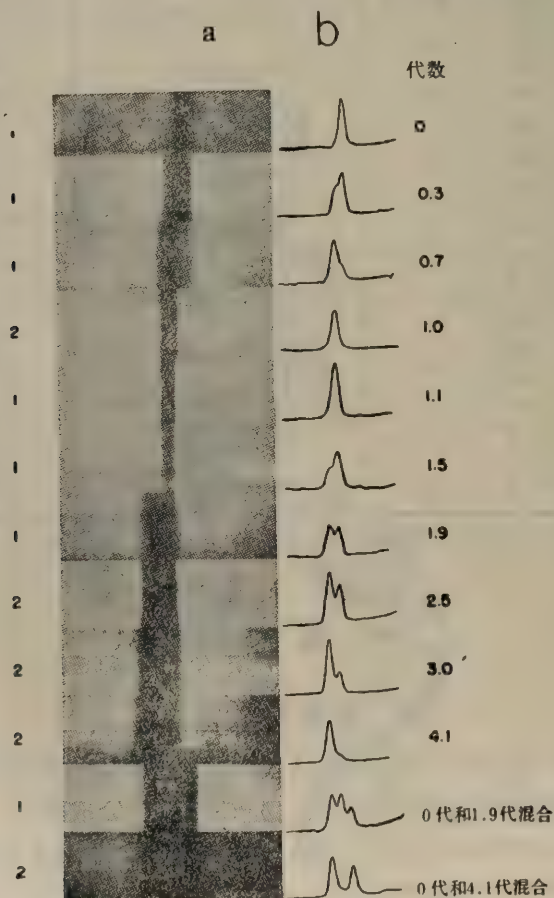


图 4-3 ^{15}N 标记 DNA 在复制时密度的变化

紫外光下可以看到一条吸收带。通常 DNA 分子的密度在 1.71g/cm^3 左右。用 ^{15}N 标记的 DNA 密度大于正常的 DNA 密度，它在离心管较低的部位形成一条区带。让这种重标记 DNA 在 ^{14}N 培养液中生长一代，我们看到新复制的 DNA 在密度稍低的地方形成一条新的区带，称为中间密度带。注意，这时在重密度区域的位置已没有 DNA 区带了。再复制一代，在 CsCl 密度梯度中出现了二条带，一条在中间密度的区域，另外一条在轻密度的区域。当然重密度的区域仍然没有 DNA 区带。随着在 ^{14}N 培养液中培养的代数增加，中密度区带的量逐渐减少，低密度区带的量逐渐增加，但是并不出现具有新的密度的区带(图 4-3)。

这个实验结果表明在体内大肠杆菌 DNA 是以半保留的方式进行复制的。在 ^{14}N 培养的第一代 DNA 中，重密度区带完全消失，排除了全保守的复制方式，因为在全保守复制方式中亲代 DNA 毫无变化转移到子代中，重密度的 DNA 区带不会消失，在以半保留方式复制的子代 DNA 中，一条链是亲代 ^{15}N 标记的重链，另一条链是含 ^{14}N 的轻链，二条链形成密度上杂交的 DNA。因此所有的子代 DNA 的密度都比亲代 DNA 的密度小，但又比正常 DNA 的密度大，形成一个中间密度的区带。如果把这种密度上杂交的分子在 ^{14}N 培养液中再培养一代，按半保留复制方式就会形成 ^{15}N^{14}N 和 ^{14}N^{14}N 二种分子，而且随着在 ^{14}N 中复制的代数增加，^{15}N^{14}N DNA 的相对量减少，而 ^{14}N^{14}N 的 DNA 的相对量增加，但不会出现其他密度的 DNA。实验结果正符合这种半保留复制的方式。

1963 年 Cairns 用放射自显影的方法阐明大肠杆菌 DNA 是一个环状分子，并以半保留方式进行复制。把大肠杆菌放在含 ^{3}H 脱氧胸苷培养液中生长二代左右，然后用溶菌酶把细胞壁消化掉，提取 DNA，十分小心地放在一张透析膜上，再把膜转移在显微镜载片上，上面覆上照相乳胶，在暗处曝光二个月。在这期间 ^{3}H 由于放射性衰变，放出 β 粒子，使乳胶曝光。显影以后，发射的 β 粒子作为黑点记录下来(图 4-4)。因为 ^{3}H 的 β 粒

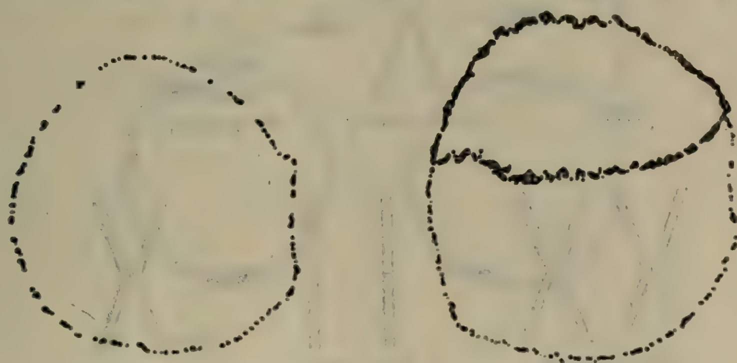


图 4-4 复制中的大肠杆菌染色体放射自显影照片

子能量很低，只能透过 1μm 距离，银粒子的位置正确地代表了 ^{3}H 在染色体中的位置。在乳胶中黑点的数目代表了 ^{3}H 在 DNA 分子中的密度。把显影好的载片放在光学显微镜底下就可以观察到染色体的形状。用这方法 Cairns 观察到大肠杆菌 DNA 是一个环状分子(图 4-4 左)。当 DNA 复制时它的中间产物是 θ 形的(图 4-4 右)，非复制的部分银粒子的密度较低，由一条"热链"和一条"冷链"构成。复制中的二条双链，其中一条双链的一条链是标记的(一条"热链"，一条"冷链")，另一条双链的二条链都标记。因此可以推断

在 DNA 复制过程中必须经过链分离阶段,复制是以半保留方式进行的。

真核生物的染色体复制也是以半保留方式进行的。1957 年 Taylor 把蚕豆苗放在含 ^{3}H 脱氧胸苷的溶液中生长,所有在细胞分裂过程中合成的 DNA 都被同位素标记了,8 个小时以后转到正常的溶液中让它继续生长,然后把细胞染色体放射自显影。从显影图上可见,在 ^{3}H 脱氧胸苷中生长一代以后,每个中期染色体的二个染色单体都是标记的。在正常介质中生长一代以后,每个中期染色体有一个染色单体是标记的,另一个是非标记的。再生长一代后,有一半染色体是非标记的,还有一半染色体中一半染色单体是标记的,另一半是非标记的(图 4-5)。

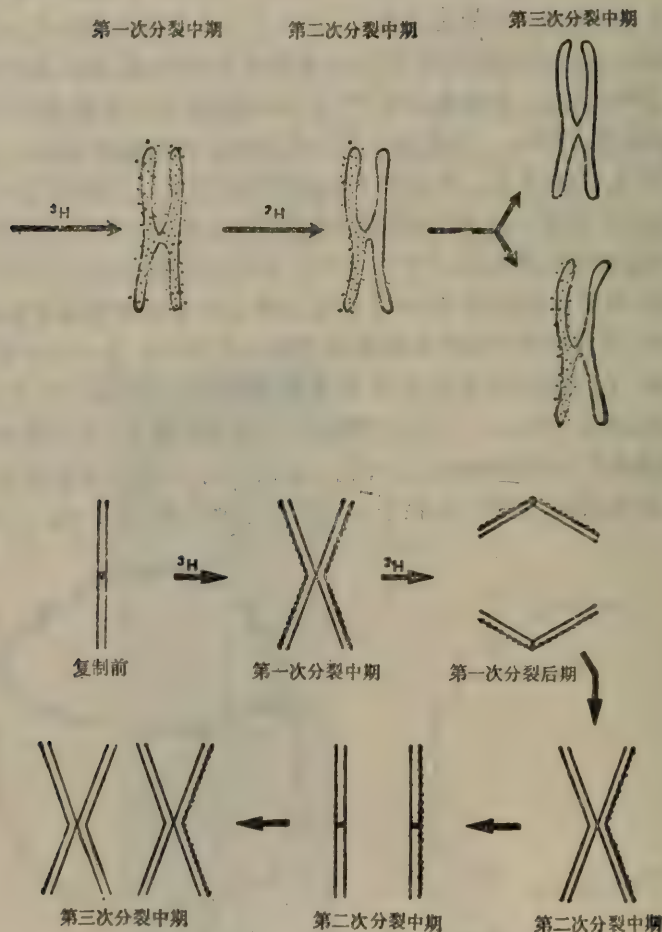


图 4-5 [^{3}H] TdR 标记的蚕豆染色体在各代中的放射自显影示意图
小点表示有放射性的部位,下图是上图的解释。

对上述实验结果最简单的解释是复制前的染色体是双元结构,在第一次分裂中间,每一个染色单体是由一条标记的和一条非标记的链组成,在正常培养液中培养一代后,每一个中期染色体将有一个染色单体中一条链是标记的,另一个染色单体完全没有标记。这个结果表明在真核生物中染色体的复制也是以半保留的方式进行的,而且表明一个染色体中只含一个非常长的 DNA 分子。

从以上的实验得到一个共同的结论，DNA 双链复制是以半保留方式进行的，无论是原核还是真核生物都是如此。

θ 方式和滚筒方式

复制时双链 DNA 解成二股链分别进行，因此这个生长点呈现叉子的形式称为复制叉。双链环状 DNA 复制的中间产物形成一个 θ 形式。这是由 Cairns 在大肠杆菌染色体的复制的中间产物放射自显影图上最早观察到的，因此这种复制方式称为 Cairns 方式。

另有一种滚筒式复制模型，这个模型可以解释噬菌体 ϕX174 复制的中间过程，特别是解释单链的子代 DNA 怎样从双链的复制型产生的。实验发现噬菌体 ϕX174 初始合成的单链产物比单位长度要长得多，而且是和亲代分子连在一起的。若 ϕX174 共价闭环的双链复制型（RFI）的正链由一内切酶切开一条链，原来的 5′ 端从双链上“剥出”，然后可能附着在膜上，留下正链的 3′OH 作为引物，以负链作为模板就可连续合成比原长的度长达几倍的正链。现在有些实验指出某些双链 DNA 的合成可能也是以滚筒的形式进行的。当双链 DNA 以这种方式合成时，正链以上述的方式“滚出”，而负链则以

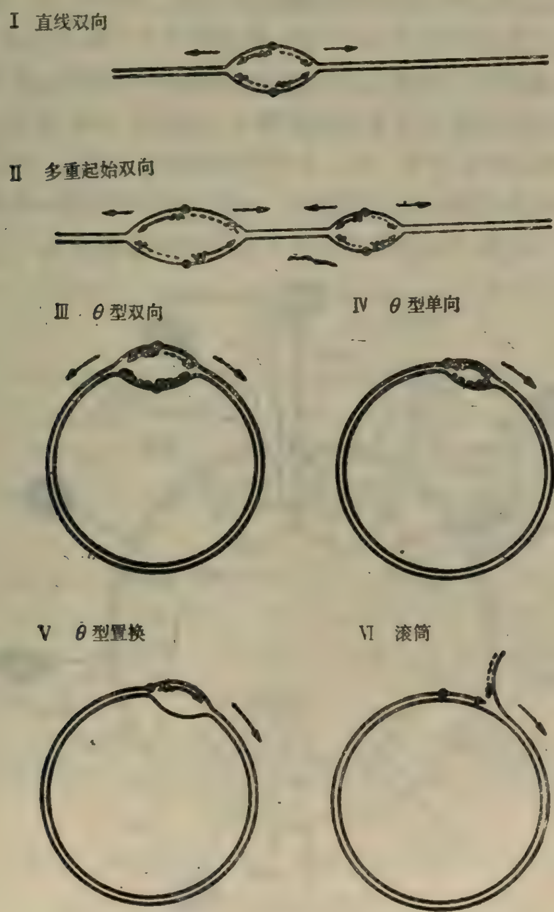


图 4-6 半保留复制的几种方式

滚出的正链为模板连续合成。在噬菌体 λ 复制的后期也可以看到这种合成方式。爪蟾卵母细胞中核糖体 RNA 基因的扩增可能也是以这种方式进行。 但是总的来说这都是一些特例，绝大多数的复制是以复制叉的形式进行的。 图 4-6 归纳了上述几种复制的机制。

复制的起始和方向

一个 DNA 分子的复制当然不是瞬时能完成的，复制过程在空间上是井然有序的复杂生化反应历程。 那么从哪一个位置首先开始复制？ 这个位置在 DNA 分子上是任意的还是特定的？

从很多体系用不同的实验手段所得出结果都证明，原核生物 DNA 复制都是在 DNA 分子特定的位置起始的，复制方向大多是双向进行的，也有一些以单向的或不对称方式进行的。

用遗传的方法可以证明大肠杆菌 DNA 复制是定点开始双向进行的（图 4-6、III）。细菌染色体复制所需要的时间相当于细胞生长一代的时间，在一个生长的群体中几乎所有的染色体都在复制中。假定每个染色体都在同样的起始点(i)开始复制，那么大部分细菌将复制 i 附近的基因，而远离 i 的基因将复制得少些。因此在一个群体中离 i 点越近的基因出现频率越高。假如复制是单向进行的，基因频率在基因图上呈单向梯度，最高和最低频率的基因应该是连锁的。如果是双向复制，基因频率将以 i 为中心出现一个双向的梯度。测定了大肠杆菌染色体上很多基因的频率，发现以 oriC 为中心双向下降，因此证明大肠杆菌 DNA 复制是起始于 oriC 基因附近双向进行的。 大肠杆菌复制起始区 oriC 位于其遗传图的 ilv 基因处（83 分附近）。一旦复制起始以后，复制叉就以双向等速前进，两个叉在距起始 180° 的 trp 基因处(33 分附近)相会(图 4-7)。

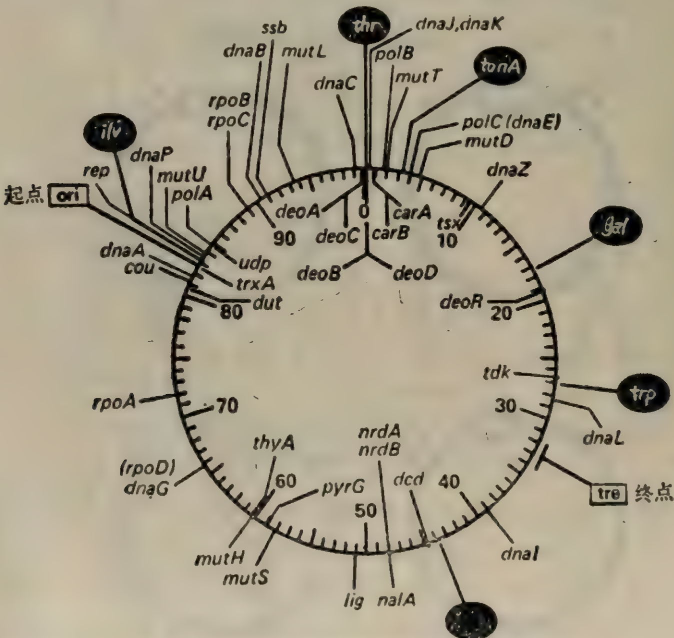


图 4-7 大肠杆菌复制起始区在遗传图上的位置

大肠杆菌和其他几种革兰氏阴性细菌起始区的顺序已经测定(图 4-8),测定结果表明,这一起始区的结构是很保守的,其中 GATC 和 TTATCCACA 出现多次。这 245 碱基对能被细菌酶系识别。

直接的放射自显影观察可以确证复制是双向的。利用大肠杆菌一个温度突变种,它在 42℃ 能使复制中的 DNA 合成完成,但不能开始新的复制循环,而在 25℃ 生长是正常的,因此有了一个简单的手段保证群体中所有的细胞都在同一时间开始复制,即称为同步生长。细菌先在含低比放射性的 ^3H 脱氧胸苷溶液中进行标记,然后在第一个复制循环即将结束、第二个循环刚开始之前几分钟转移到高比放射性的 ^3H 脱氧胸苷溶液培养一段不长的时间。假如复制是单向的,起始点和终止点应该是邻近的,如果是双向的,则二个区域应该是分离的,实验结果表明复制是双向进行的(图 4-9)。

以上是原核生物中的一些实验结果,在真核生物中,实验表明,复制也是从特定的位置起始,以双向延伸的方式进行,在 SV40DNA 复制时,随着时间延长,增加标记的比强度,把复制完成的 DNA 分子分离出来,用限制性内切酶消化,测定每个片段的比放射性强度,并按物理图谱排列起来,发现放射性强度呈梯度分布,最小的放射性区域是复制的起始点,二边的放射性对称地增强。用限制性内切酶把 SV40 DNA 切成线状,作为 DNA 分子内位置的标志,用电镜也可观察到复制眼和它的延伸过程,因此 SV40 DNA 的复制也是从定点开始,双向进行的。

用哺乳动物细胞做材料,从放射自显影图上也可看到复制是双向的,而且在第一个复制周期所用的起始点看来也用在第二个复制周期中,这些材料都说明 DNA 复制是在特定的区域起始,向两个方向等速进行的。

大多数生物体内 DNA 的复制都是以这种方式进行的,但是也有一些例外的情况,例如在枯草杆菌中,复制是从 DNA 上特定的点开始的,但是二个复制叉的移动是不对称的。一个复制叉仅移动染色体的五分之一的距离,然后停下来等另一个复制叉走完五分之四的距离。

用限制性内切酶的切点作为参考点,用电镜观察分离质粒 R6K 的复制中间产物,发现在复制叉移动不太远的复制中间体中,复制叉离开参考点的距离是恒定的,说明复制的早期是单向进行。然而复制得更远的中间产物显示出第一个复制叉在离起始点 20% 左右的位置停下来,而从反方向开始形成第二个复制叉,因此质粒 R6K 的复制是不对称进行的。

质粒 ColEI 的复制完全是单向进行的。在 ColEI DNA 中限制性内切酶 EcoRl 只有一个切点,这个切点可以作为参考点。从体内或体外合成系统分离到的复制中间体其一个复制叉距离参考点约 17—18% 左右,而另一个复制叉的距离是可变的,因此说明 ColEI 的复制起点是固定的,但只向一个方向移动。

线粒体 DNA 的复制也是不对称进行的,DNA 双链中一条链先复制,另一条链保持单链,电镜中呈 D 环形状,待一条链复制完毕,另一条链才开始复制(见图 4-6、V、Q 型置换方式)。

从已研究过的 DNA 复制形式来看,复制是在 DNA 分子中一个特定的位点开始的,至今还没有例外。复制叉移动的方向和速度是多种多样的,但主要是双向等速方式移动的。

```
                         BglII          BglII       50                                                    BamHI  100
相同顺序:         ACAAGATCTnTTTATTTAGAGATCTGTTnTATTGTGATCTCTTATTAGGATCGnnntnnnnTGTGGATAAgnngGATCCnnn
                                                                                 ACACCTATT
大肠杆菌         : ..      A        .      .  C    ..           CACTGCCC        CAAG.....GGC
鼠伤寒沙门氏菌    : ..      T        .      .  C    ..           CGCCAGGC        CCCG.....TGT
Enterobacter aerogenes: .. -        .      .  C    ..           ACTCTCTA        GTCG.....ACG
肺炎杆菌         : ..      -        .      .  T    ..           GCTTGTCT        GTCA.....GCG
Erwinia carotovora: GA     T        A      C  -    A.           TCGTGTTG        GTGATTATTCAT

101                                              150  AvaII                                          200
nTTtAAGATCAAnnnnnTggnAAGGATCnctAnCTGTGAATGATCGGTGATCcTGGnCcGTATAAGCTGGGATCAnAATGngCGnTTATACACAgctCAA
T. T A    CAACC GGA...   AT..A                  . ..A .             .G  . AG..G    .    A.T
AA A A    TGCGT GGA...   AC..G                  G ..T .             .A  C GGTAC    .    A.T
A. T A    ACGCT AAG...   ACA.T                  . .TT .             .G  . AA..G    G    G.T
G. T G    CCGTT AAG...   GC.TT                  . ..T .             .A  . AA..G    .    G.A
A. A A    GAGAA GGCGTT   CT..C                  . A.C A             TT  . TG..T    .    GGA

201                                         HindIII 250
AAAncgnACaaCGGTTaTTcTTTGGATAACTACCGGTTGATCCAAGCTTTtnAnCAgAGTTATCCACAntnGAnnGcnn-GAT : 相同顺序
                  AA^A_CCTATT
 .CTGA. AA.A  G ..  .     . .  ... .......CCTGA  G..     GTA. TC.CAC-.  : 大肠杆菌
 .GTGA. AA..  A ..  .     . .  ... ........CC.C G.T     ATG. TC.CAC-.  : 鼠伤寒沙门氏菌
 .GCAT. TC..  A ..  .     . .  ... ........TG.G GG.     GAA. AGCTGCG.  : Enterobacter aerogenes
 TTCAGG AA..  A A.  .     . .  ... ........TA.G T..     GAA. AA.TAT-.  : 肺炎杆菌
 .ACGC. TCG.  G .A  G     G C  TTA ACCAGAA.TA.G T..     TTCA CT.CCG-A  : Erwinia carotovora
```

图 4-8　五种革兰氏阴性菌复制起始区的结构

五种菌在特定位置上出现完全相同的以大楷大号字母表示;四种菌相同的以大楷小号表示;三种相同的以小楷小号字母表示;二种以下则以 n 表示。

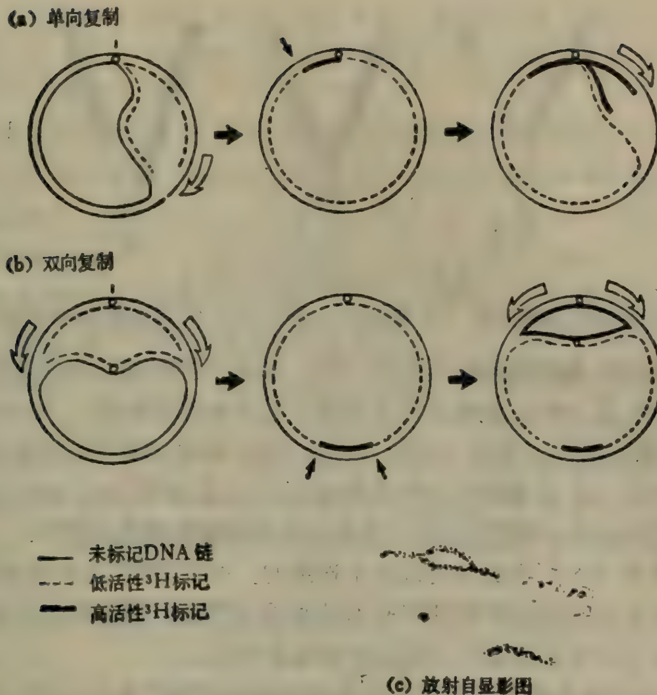


图 4-9 双向复制的放射自显影证明

半不连续复制 (Semidiscontinuous replication)

上面我们讨论了 DNA 复制的一些基本机制，即 DNA 是以半保留方式进行复制的，复制的起始点处在 DNA 分子的特定的位置上，复制叉移动的方式是多样的，但主要以对称方式移动的。现在我们讨论复制生长点附近一些情况。

因为 DNA 双螺旋的二条链是反向平行的，因此在复制叉附近解开的 DNA 二条链一条是 5′→3′ 方向，另一条是 3′→5′ 方向。由于模板的极性不同，是不是一种 DNA 聚合酶能以 3′→5′ 方向延伸，另一种聚合酶可以 5′→3′ 方向延伸呢？至今发现的体内所有 DNA 聚合酶都只能以 5′→3′ 延伸，而且以后我们将讨论到，由于复制真实性的要求，也只能以 5′→3′ 延伸。如何解决复制过程中二条不同极性的链同时延伸这个问题呢？

1968 年冈崎 (Okazaki) 发现大肠杆菌用 ^{3}H 脱氧胸苷短时间标记，从中得到的 DNA 变性后用超离心方法可以得到许多 8 S—10S 的标记小片段，这种小片段现在常称为冈崎片段。标记时间延长，冈崎片段可转变为成熟的 DNA 链，因此它们必然是复制过程的中间产物。按照这个结果，冈崎认为 DNA 合成至少在一条链上是不连续进行的。以复制叉向前移动的方向为标准，一条模板链是 3′→5′ 方向，称为前导链 (leading strand)。原则上在前导链上，DNA 合成能以 5′→3′ 方向（指新合成的 DNA 链的极性）连续合成。另一条模板链上新生的 DNA 也是从 5′→3′ 合成，但是相反于复制叉前进的方向，这一条链称为后随链 (lagging strand)，随着复制叉的移动，后随链上很多不连续的 DNA 片段可以连接起来成为一条完整的 DNA 链(图 4-10)。

还有一些实验证明，复制过程中出现冈崎片段，例如在 DNA 连接酶的突变株中积

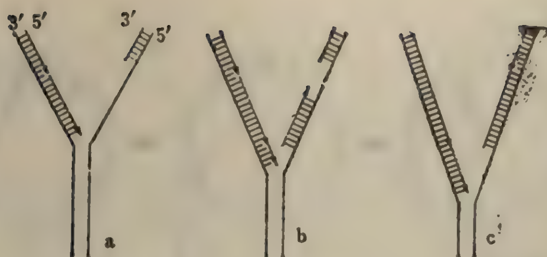


图 4-10 半不连续复制的示意图

a）前导链上连续合成； b）后随链上不连续合成； c）连接成完整复制链。

累冈崎片段，而 DNA 连接酶的作用是在 DNA 模板上把相邻二个片段连接起来。因此这个实验结果表明，在 DNA 合成过程中确实经过片段连接阶段。另一个证据是从电镜观察得到的，电镜技术可以区别 DNA 双链和单链，因此能直接看到复制叉附近的一些情况，例如在噬菌体 T7 的复制中间物的电镜照片上，在复制叉的一边是双链，另一边是单链，或一段很短的双链，这种情况可以用半不连续复制的模式来解释。

理论上，前导链 DNA 合成也可以是不连续的。由于 DNA 的合成酶系不易从 DNA 的模板上解离，因此实际上前导链 DNA 合成总是连续的。不连续的前导链 DNA 合成如果也被找到，也只能看作是这个一般规律的例外情况。

这种前导链的连续复制和后随链的不连续复制在生物界是有普遍性的，称为双螺旋的半不连续复制。有很多因素，例如模板的损伤，复制蛋白或 dNTP 供应不平衡等都可以影响前导链的连续复制，引起前导链从另一新点起始。

RNA 引 物

目前已知的 DNA 聚合酶都只能延长已存在 DNA 链，而不能重新合成 DNA 新链，这就造成了一个理解复制机制的重要障碍。一个新的 DNA 复制怎样才能开始？现在知道，在 DNA 模板上先合成一段 RNA 引物，DNA 聚合酶从 RNA 引物 3′ 端开始合成新的 DNA 链。这个问题开始是从观察利福霉素对噬菌体 M13 复制的影响得到启发的，因为利福霉素是 RNA 聚合酶的抑制剂(见第五章)。为了证实 RNA 是 DNA 合成的引物，用了不少实验方法，从各个角度获得证据：

（1）标记转移：分离[α-^{32}P] dNTP 参入的 DNA 及进行碱水解，如果 RNA 和 DNA 是共价连接的，^{32}P 应出现在核糖核苷酸中。在一些体外合成系统中，如噬菌体 M13，ϕX 174 的单链 DNA 复制成双链 DNA 时发现是这种情形。

（2）共沉降：在 DNA-RNA 变性的条件下观察标记的 RNA 是否与 DNA 一起沉降。用噬菌体 G4 做的实验结果证明 RNA 和 DNA 是共价连在一起的。

（3）密度移动：在用碱水解或用 RNase 处理前后，观察 DNA 在 CsCl 密度梯度离心中变化的情况。因为 RNA 的密度比 DNA 大，因此带有 RNA 的 DNA 分子密度也要大些。体内合成的大肠杆菌 DNA 就显示这种密度移动。

（4）在氯霉素存在的条件下复制出来的质粒 ColEI DNA 分子包含有 RNA 分子。

（5）RNA 共价连接到 DNA 末端，妨碍了 DNA 末端用多核苷酸激酶磷酸化，而 RNA 对碱是不稳定的，除去 RNA 后产生新的 DNA 的 5′OH，再标记上去的磷酸是对

碱稳定的。从大肠杆菌中分离出来的冈崎片段具有这个特性。

(6) 分离出的 RNA 引物经常只能杂交到双链 DNA 的一条链上，在噬菌体 λ 和 T7、T4 复制中可看到这种例子。

以上这些实验结果都有利于说明 RNA 引物共价连接在新生的 DNA 链上。但是对上述的证据也可从不同的角度提出反对意见。对密度迁移实验的结果现在认为是不可靠的。用碱处理在 DNA 分子上产生可以磷酸化的 5′OH 这个证据也不可靠。因为在 DNA 修复系统脱嘧啶的 DNA 链对碱也不稳定，会产生 5′OH 末端。企图直接证明 RNA 和 DNA 共价相连，困难在于引物 RNA 的量极微，而体内其他的 RNA 极丰富，混杂痕迹量的 RNA 就可能使实验结果不易解释。

关于冈崎片段带有 RNA 引物最过硬的证明是用多瘤病毒完成的。在体外核系统中合成的冈崎片段 5′ 端是一个约 10 核苷酸长的以 5′ 三磷酸为结尾的 RNA。RNA 的顺序近来已初步了解。表 4-2 列出了几种生物的 RNA 引物长度及其主要结构。RNA 引物的长度不是恒定不变的。在动物细胞中引物长度接近 10 核苷酸，第一个核苷酸常是 ATP，引物 RNA 的起始和终端的位置受解链酶 (helicase)、引发酶 (primase)、模板顺

表 4-2 DNA 复制时 RNA 引物的长度和结构

基因组	RNA 引物长度	主要结构
T7	1—5	$pppApCpC/A(pN)^{*}_{1-2}$
T4	1—5	$pppApC(pN)_3$
大肠杆菌	1—3	未测定
φX174	1—5	pppA……
SV40	~10	$pppA/G\ (pN)_9$
海胆	1—8	$pppA/G\ (pN)_7$

* C/A 表示此位置是 C 或 A，N：A 和 C 较常见。

序及其二级结构的影响。

引　发

DNA 复制过程中需要合成 RNA 引物，这可以说是合成后随链的普遍规律。RNA 引物的合成称为引发。引发是一个十分复杂的过程。

现在知道大肠杆菌 DNA 的复制需要多种蛋白质和酶的协同作用，其中的不少蛋白质都与引发有关(见表 4-1)。

由于大肠杆菌的引发机制过于复杂，人们首先研究了大肠杆菌噬菌体的引发过程。在大肠杆菌中，噬菌体 M13 DNA 可以利用细菌本身的 RNA 聚合酶，首先在 M13 单链 DNA 上合成一小段 RNA 引物，然后在 DNA 聚合酶的作用下，以 M13 为模板延伸新生的 DNA 链，图 4-11 表示 M13 的全部复制过程。在有利福平时，这种单链 M13DNA 变为双链 DNA 的反应就完全被抑制了，因为利福平可以抑制大肠杆菌 RNA 聚合酶的活力(见第五章)。

但是，在一种大肠杆菌的突变株中，M13 DNA 的合成不受利福平的抑制。此外，单链 φX174 DNA 变为双链的 RF 型也完全不受利福平的影响，因此可能存在一种以前未

认识到的 RNA 合成系统。追踪的结果导致了引发体（primosome）的发现。已经有证据表明大肠杆菌、枯草杆菌、SV40 和淋巴细胞内 DNA 复制时都要靠它来引发。

图 4-12 是 ϕX174 单链 DNA 转变为双链复制型 DNA 的全过程，包括：I、识别，II、组装，III、移动，IV、引发，V、延伸和 VI 切除(缺口填满和连接)，其中引发前体(preprimosome) 由 n、n′、n″、dna B、dna C 和 i 六种蛋白质组成，这六种蛋白质都已

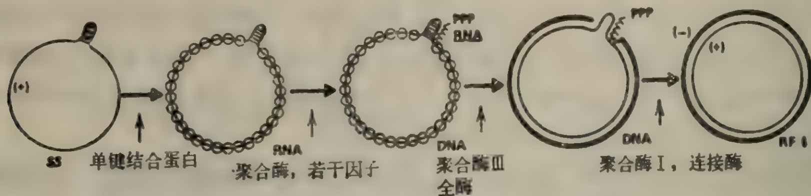


图 4-11　M13 由单链 DNA 变为双链时的必要过程

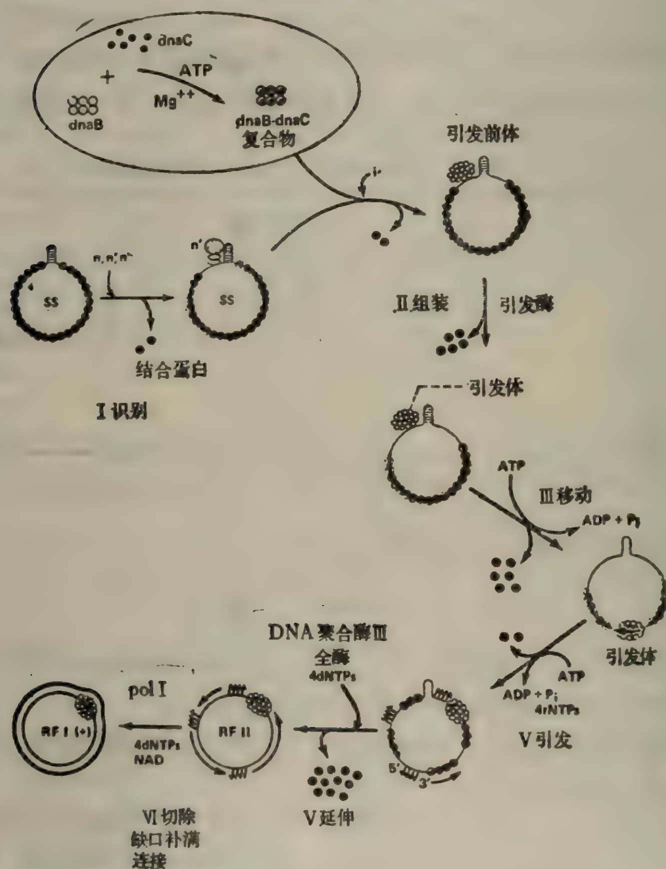


图 4-12　ϕX174 DNA 由单链转变为双链时的必要步骤

提纯。只有把六种蛋白质合在一起并与引发酶（primase）进一步组装成引发体时才能发挥其功效。

引发体可以像火车头一样在后随链上向分叉的方向前进，并在模板上断断续续地引发生成后随链的引物 RNA“短”链（图 4-13)。dna B 蛋白的作用就在于识别大肠杆菌 DNA 的起始位置，结合在两条链上并以 5′→3′ 方向运动(图 4-14)。

链的延伸

前面讨论 DNA 复制的引发机制时已经讨论到DNA 链的延伸问题(图 4-12—图 4-14)。链的延伸是在 DNA 聚合酶催化下进行的,它可以在引物的 3′-OH 上一个一个地按模板要求延伸新生链。在大肠杆菌中发现有三个 DNA 聚合酶,按发现先后分别叫 Pol I、Pol II 和 Pol III。在体内,Pol III 是延伸 DNA 的主要的酶,这个酶的缺陷株

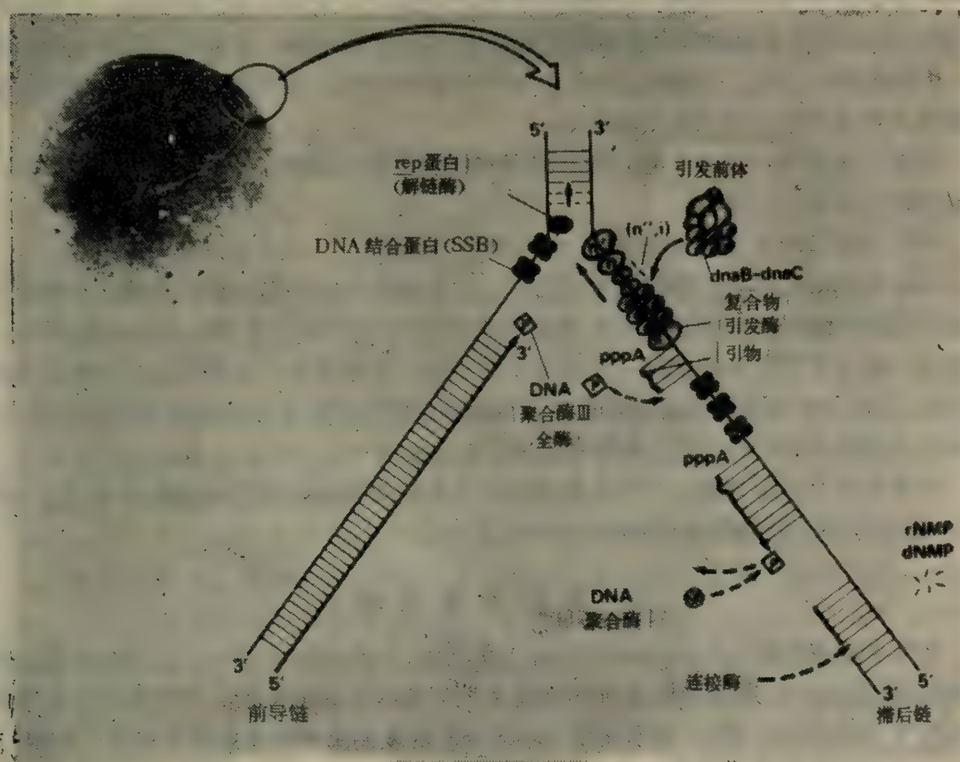


图 4-13 大肠杆菌双向复制的示意图

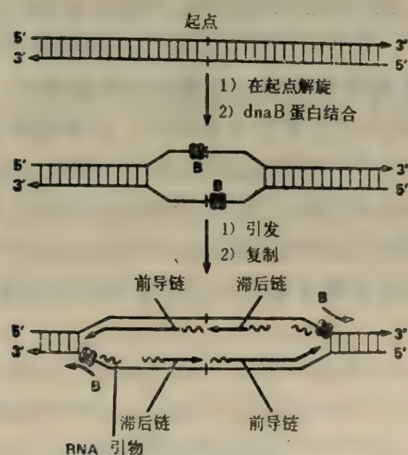


图 4-14 复制时两个分叉的双向运动

是致死的，而 Pol I 和 Pol II 的突变株则是能存活的。按 DNA 复制的速率计算，也只有 Pol III 能满足体内复制的需要。

关于 DNA 聚合物的详细性质见第二节。

链 的 终 止

DNA 链在延长阶段结束后，就会迅速地受到酶的作用，切除引物 RNA。已经知道，RNase H 能识别 DNA-RNA 杂交链并水解杂交链中的 RNA。但 RNase H 不能完全去除 RNA，很可能在生物体内引物的除去要由 DNA 聚合酶 I 来执行，因为这个酶具有 5—3′ 外切酶的功能。引物除去后所留下的空缺靠 DNA 聚合酶 I 利用四种脱氧核苷三磷酸为原料按模板的要求将空缺补满。

空缺补满后再靠 DNA 连接酶将相邻二个核苷酸以磷酸二酯键连接起来，DNA 的复制到此可以说告一段落。

环状双链 DNA 双向复制中有一个问题：复制在什么地方终止?

一般来说链的终止不需要特定的信号，也不需要特殊的蛋白质来参与。例如噬菌体 λ 就没有特定的终止顺序。λDNA 按 θ 方式复制时，它的终止点和起始点相差 180°。如果删去一些非必需区域或者插入一些无关的片段，这就改变了复制起始点和终止点的相对距离，如此一个突变株的终止点仍在起始点的 180° 位置，而不是在原来的终止点上。这个实验表明在 λDNA 分子上没有特定的终止点。

但是在不对称复制时，一个复制叉怎么会在复制的中途停下来还很不清楚。线状 DNA 复制是如何结束的也有待进一步研究。

复 制 的 忠 实 性

在双螺旋模型提出和 DNA 聚合酶 I 发现后，很多人认为 DNA 复制的问题已解决了，现在看来并非如此，DNA 复制机制比人们想象的要复杂得多。生物体为什么用这样的机制来复制 DNA 呢？ 能不能用一种 3′→5′ 的复制酶解决双螺旋反向平行问题？ 在生物体内这个重要的过程中为什么表现得如此呆板，在能量上如此浪费呢？ 其实这一套复制机制可能有利于复制的调控和复制忠实性的维持，对前一点我们了解不多，对后一点人们已有初步的认识，DNA 复制的高度忠实性要求生物体具有如此的机制。

大肠杆菌复制 10^9—10^{10} 碱基对仅有一个误差。如前所述，DNA 聚合酶要求有引物存在，而且引物的 3′OH 末端必须和模板正确配对。如果模板引物带有一个错配的末端，那么聚合酶将应用它的 3′→5′ 外切酶活力切去错配的碱基，这个反应一直进行到出现正确碱基配对的末端为止。如果聚合酶有选择正确碱基的能力，同时又有独立的校对功能，就可以保证复制的高度忠实性。从化学角度估计碱基误配的机率不可能小于 10^{-5}，在这种情况下，至少要有一个校对步骤才能使误差水平降到可以耐受的水平（$10^{-5} \times 10^{-5} = 10^{-10}$）。

按照这个观点，DNA 复制机制中的一些现象就都可以得到合理的解释。

(1) 为什么没有一个 DNA 聚合酶具有开始合成新的核酸链的能力而必须要用一个 3′ 端配对的引物来开始一个新的链？这是因为自我校对的功能必须要使 DNA 聚合酶能识别末端是否正确配对，而另一方面如果要求酶在没有引物的情况下开始新链合成，而又

不失去对碱基配对和不配对的 3′OH 的鉴别，这是自相矛盾的。为了保证高度的真实性，生物体就选择了不能重新开始新链合成的 DNA 聚合酶来复制 DNA。

（2）为什么生物体内找不到 3′→5′ 方向的聚合酶？如果 DNA 链是 3′→5′ 生长，则生长中的链 5′ 端带有活化三磷酸，这使得聚合的误差不能得到校正，因为切去误配的 5′ 端就失去进一步合成所必须的高能键，除非生物体内有另一套酶系统再磷酸化 5′端，而这在能量利用上就不经济了。

（3）为什么引物都是用 RNA？因为 RNA 引物都是重新开始合成的，它错配的可能性大，用 RNA 做引物，它自动标志这一般核苷酸是必需删去的坏拷贝，这种明显的区别在进化上是有优越性的。

小　结

综上所述，DNA 的复制遵循以下基本法则：

（1）复制是半保留的。

（2）复制起始于细菌或病毒的特定位置，真核生物有多处起始点。因此大肠杆菌复制一代约需 20 分钟，而染色体 DNA 比大肠杆菌大 40 倍的果蝇只要 3 分钟就可以完成其 DNA 的复制。

（3）复制可以是单向的，也可以是双向的。双向复制的速度不一定相同。

（4）复制的速度取决于起始。复制叉前进的速度是比较恒定的，原核生物每分钟约为 10^5 碱基对，真核生物在 5×10^2—5×10^3 碱基对。

（5）复制时每条链都由 5′ 端向 3′ 端延长，每次增加一个核苷酸。

（6）复制是半不连续的。在前导链上是连续的，在后随链上是不连续的。在动物细胞中每个冈崎片段长约 100—200 个核苷酸，在原核生物中每小段长约 1,000—2,000 个核苷酸。这些小片段再由 DNA 连接酶连接到生长中的长链上去。

（7）新生 DNA 靠 RNA 来起始。这一小段作为引物的 RNA 在以后被酶切除。缺口由 DNA 补满并与生长中的 DNA 连在一起。

（8）双向复制有时在一固定的位置结束。

（9）即使在一个单细胞中，复制也可以采用不同的方式。

二、DNA 聚合酶和 DNA 连接酶

DNA 聚合酶是指以脱氧核苷三磷酸作为底物催化合成 DNA 的一类酶，从广泛不同的生物，包括细菌，植物，动物中都发现这种酶，所有 DNA 聚合酶作用方式基本相似，它们催化脱氧核苷三磷酸加到复制中 DNA 链的 3′ 羟基末端。合成方向是从 5′→3′，添加的脱氧核苷酸的种类由模板 DNA 决定。 因此 DNA 聚合酶需要的条件包括四种脱氧核苷三磷酸，镁离子，模板 DNA 和引物。还没有发现一种不需引物能重新合成 DNA 链的酶。

大肠杆菌 DNA 聚合酶 I

大肠杆菌 DNA 聚合酶 I 是 1956 年 Kornberg 及其合作者发现的第一个 DNA 聚

合酶,故也称 Kornberg 酶。在各种 DNA 聚合酶中,这个酶是研究得最透彻的,它的一些特点代表了 DNA 聚合酶的基本特点,以下对这个酶作比较详细的讨论。

当存在 ATP、Mg^{2+}时,大肠杆菌的无细胞提取液能使 ^{14}C 脱氧胸苷变成酸沉淀的产物,此酸沉淀物用脱氧核糖核酸酶作用之后变成酸不沉淀的,说明这个产物具有 DNA 的性质。以后用标记的 dTTP 作为底物,以酸沉淀产物的生成作为指标,已把 DNA 聚合酶 I 纯化到相当均一的水平。

用比较纯的酶做试验发现,为了得到比较高的参入水平,除了变性的 DNA 模板和 Mg^{2+} 之外,必须有四种脱氧核苷三磷酸同时存在,缺少任何一种,参入的水平将大大降低. 脱氧核苷二磷酸不能代替三磷酸,用纯的酶制剂可以计算出 DNA 的净合成大约增加 10—20 倍,因此 90—95% 的 DNA 是新合成的。

DNA 聚合酶 I 还包含一些其他酶活性,它们是整个酶分子的组成部分。其中二个外切酶,一个 5′→3′ 外切酶,一个 3′→5′ 外切酶,后面将讨论到这二个酶活性是 DNA 复制过程中所必需的。

DNA 聚合酶 I 是 DNA 指导的 DNA 聚合酶。 但是新合成的 DNA 是否确实是在模板指导下进行的?下列证据给予明确的回答:

(1) 最明确的证据是这个酶能忠实地复制噬菌体 ϕX174DNA, 这个合成产物用多核苷酸连接酶环化以后具有生物感染活力。ϕX174DNA 是由 5386 个核苷酸构成的单链 DNA。噬菌体含有的 DNA 称正链 DNA,它的互补链称负链 DNA。正链 DNA 同大肠杆菌 DNA 聚合酶、DNA 连接酶、四种脱氧核苷三磷酸、NAD、Mg^{2+} 以及煮沸过的大肠杆菌提取液(含寡核苷酸,可能提供必须的引物)一起保温。产物是双链环状 DNA,它的性质同天然的复制型(RF)完全一样。如果用 5-溴脱氧尿苷 (BrdU) 三磷酸代替 dTTP,用 α-^{32}P 代替非标记的 dATP 合成的 RF 分子中的负链(图 4-15A)。因为含 Brd U 的负链密度比正链要重些,并带有同位素标记。在每个 RF 分子中用 DNase 引入一个切口,再热变性以后,可把正链和负链用密度梯度离心方法分开。完全新合成的负链不含天然的正链 DNA (图 4-15B)。这个 DNA 能感染原生质体,复制出新的病毒颗粒。现在再用完全是人工合成的带 Brd U 的负链 DNA 为模板,用正常的四种 dNTP 合成 RF 分子。分开正链负链之后,这样获得的正链具有天然 ϕX174DNA 一样的感染力。而且理化性质也完全同天然的一样(图 4-15C)。 这个结果表明 DNA 聚合酶 I 能在模板 DNA 指导下忠实地合成新的 DNA。

(2) 用 DNA 聚合酶 I 合成的新 DNA 样品,在化学上或物理化学上的性质都同模板 DNA 一样,具有天然 DNA 一样的双螺旋结构和同样的分子融点,也符合 A = T、G = C 的 Chargaff 规则。用碱基比例相差很大的 DNA 作为模板,新合成 DNA 的碱基比例也同所用的模板完全一样。如果用合成的多聚 d(A-T) 作为模板,即使反应混合物中有四种 dNTP,也只有 dATP 和 dTTP 参入到产物中去。

(3) 用毗邻分析的方法可以证明新合成的 DNA 的顺序同模板是完全互补的。DNA 聚合酶同模板和所有四种脱氧核糖三磷酸一起保温,其中一种,譬如 dATP 是 [α-^{32}P] 标记的,在合成反应中,这个标记的磷酸同它左边相邻的核苷酸(N)结合。 用脾磷酸二酯酶和小球菌 DNase 完全水解以后产生脱氧核苷 3′ 磷酸,^{32}P 就转移到 N_p 上(N_p 是四种 3′-核苷酸中的任何一种)。分析四种标记的核苷酸的含量,就可以推断 A 的左邻d(ApA)、

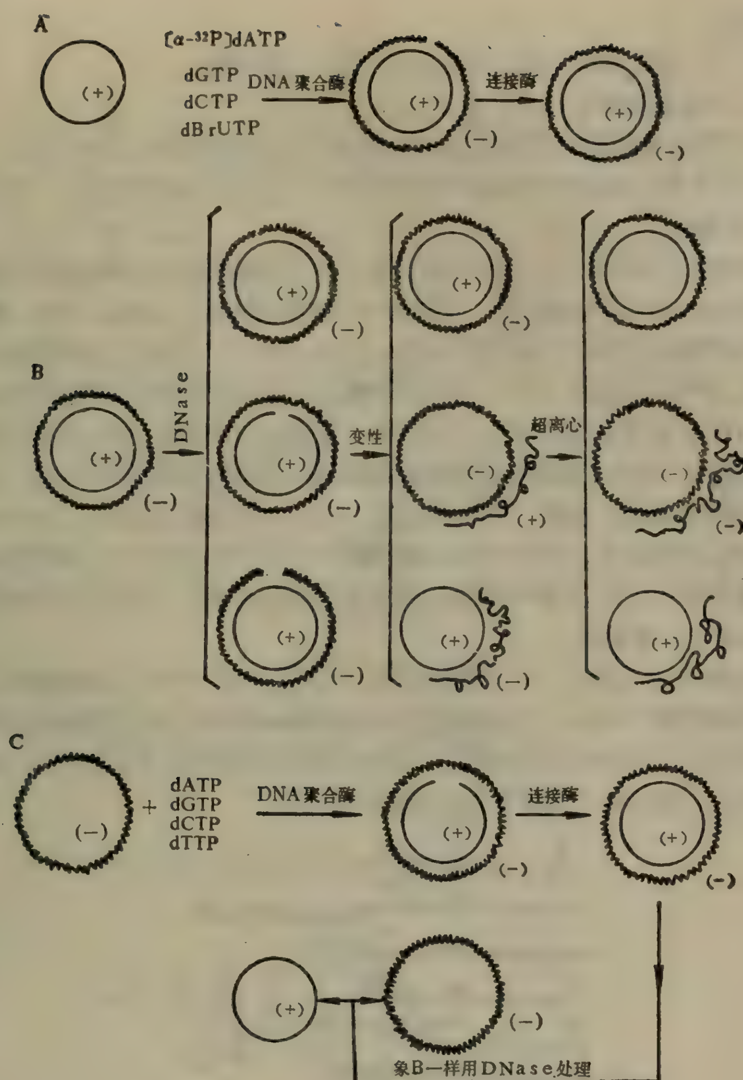


图 4-15 DNA 聚合酶的实验检查

d(CpA)、d(TpA) 和 d(GpA) 在整个分子中的频率， 同样也可推断其他各种二核苷酸的相对频率。 用这个方法可以证实新合成的 DNA 的顺序同模板 DNA 的顺序是互补的。

纯的 DNA 聚合酶 I 由分子量 109,000 的一条肽链构成，每个大肠杆菌细胞内大约含 400 个酶分子。用枯草杆菌蛋白酶处理这个酶，获得两个片段，大的片段分子量 76,000，保留聚合酶和 3′→5′ 外切酶活力(被叫作 Klenow 片段，广泛使用于 DNA 序列分析，见第二章。)。小的片段分子量 34,000，具有 5′→3′ 外切酶的活力。

DNA 聚合酶 I 具有多种酶活力。

(1) 在引物的 3′-OH 末端，以脱氧核苷三磷酸为底物在模板指导下逐个加上脱氧核苷酸，每个酶每分钟约添加 1,000 个单核苷酸。

(2) 在没有脱氧核苷三磷酸底物时，能从 3′-OH 开始以 3′→5′ 方向水解 DNA，产

生 5′ 单核苷酸。

(3) 从 5′→3′ 方向水解 DNA 链,主要产生 5′ 单核苷酸。

(4) 从 3′ 末端焦磷酸解 DNA 分子。

(5) 催化脱氧核糖核苷三磷酸的末端焦磷酸同无机焦磷酸的交换反应。

后二个反应是聚合反应的逆反应。在体内这二个反应没有重要意义。因为它们需要高浓度的无机焦磷酸。

DNA 聚合酶 I 具有 3′→5′ 的外切酶活性，这个外切酶能识别和消除不配对的引物末端。当模板和引物的末端错配时，即使存在四种脱氧核苷三磷酸，DNA 聚合酶也是首先除去了错配的碱基,然后才进行聚合反应,因此 3′→5′ 外切酶活力能够校正聚合过程中的碱基错配。DNA 聚合酶 I 的 3′→5′ 外切酶活力是聚合酶的必要组成部分。如果没有这种活性，DNA 复制的忠实性将大大降低。噬菌体 T4 的一种突变株,变异率很高,逆转率也高,这说明 DNA 复制的忠实性低,错误多。从这种突变株得到的 T4 DNA 聚合酶, 3′→5′ 的外切酶活力也低，因此降低了 DNA 的复制的忠实性。另一种 T4 突变株,具有很高的抗变异能力,它的 DNA 聚合酶中 3′→5′ 外切酶活力比野生型要高得多,DNA 复制的真实性十分好,变异率很低。这个实验结果表明，3′→5′ 外切酶的校对功能对 DNA 复制是十分重要的。

DNA 聚合酶 I 还具有 5′→3′ 外切酶活性，它能作用于 DNA 的碱基配对部分，从 5′ 端释放寡核苷酸(图 4-16)。它也能跳过几个核苷酸起作用,所以它能除去紫外光照形成的胸腺嘧啶双体。在 DNA 损伤修复中它也可能起重要作用（图 4-17)。除去冈崎片段 5′ 端的 RNA 引物可能也有赖于这个外切酶的活性。

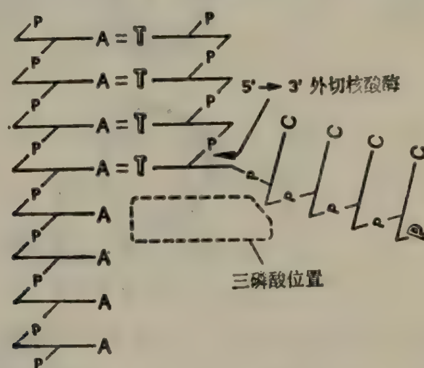


图 4-16 DNA 聚合酶 I 的 5′→3′ 外切活性

虽然 DNA 聚合酶 I 能延伸 DNA 链,图 4-18 表示 ϕX174 或 M13 可以被 Pol I 从单链环状生成复制形产物，但是，DNA 聚合酶 I 不是复制大肠杆菌染色体的主要聚合酶,因为:

(1) 纯的酶催化核苷酸参入的速率约每分钟 1,000 个核苷酸，而体内实际的合成速率要比这个速率高二个数量级。

(2) 大肠杆菌有一种突变株,从中分离出来的 DNA 聚合酶 I 的活性是正常的，但染色体复制是不正常的。

(3) 说明 DNA 聚合酶 I 不是大肠杆菌主要的复制酶的最有力的证据是分离到一个

图 4-17 DNA 聚合酶 I 删除胸腺嘧啶双体

突变株 ($PolA^-$)，它的 DNA 聚合酶 I 的活力仅仅是野生型的 1%，但它的复制是正常的。这个突变株减低了冈崎片段连接的活力，增加了对紫外光、烷化剂和离子辐射的敏感性。这意味着 DNA 聚合酶 I 在大肠杆菌 DNA 复制中不是主要的合成酶，而是在除去 RNA 引物和损伤修复中起作用。

大肠杆菌 DNA 聚合酶 II

纯化的 DNA 聚合酶 II 分子量 90,000—120,000，由一条肽链构成。它也是从 5′→3′方向合成 DNA，达到最大活力需要四种脱氧核苷三磷酸，Mg^{2+}，NH_4^+ 和一个带 50—200 碱基对缺口的双链 DNA，当缺口加大时，反应速度降低，但加入 DNA 结合蛋白时能恢复到原来的活力。这个酶也需要带 3′-OH 末端的引物，具有 3′→5′ 的外切酶活性，但没有 5′→3′ 外切酶活性。它的合成速率比体内的速率要低得多，带有这个酶缺陷的大肠杆菌突变株染色体的复制各方面都正常，它在体内的功能可能和 DNA 聚合酶 I 类

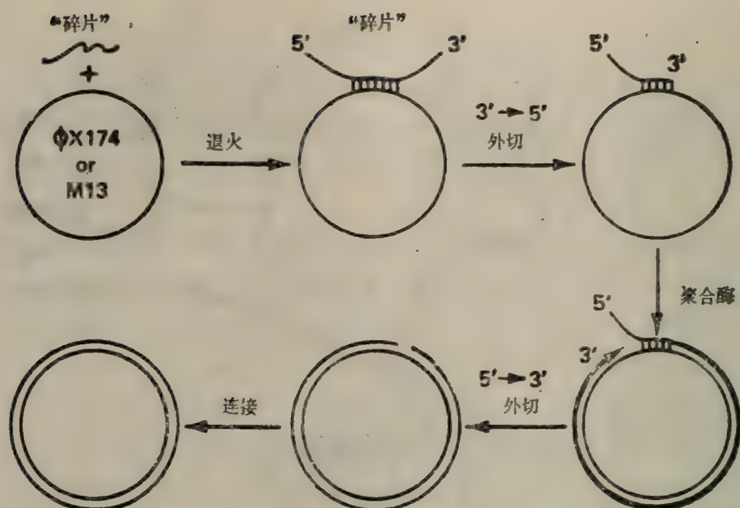


图 4-18 DNA 聚合酶 I 在引物上延伸 DNA 链
(依赖于该酶的三种酶活性)

似。

大肠杆菌 DNA 聚合酶 III 全酶

带有 DNA 聚合酶 III 的温度敏感突变(pol C) 的大肠杆菌在限制温度是不能存活的,从这种菌株得到的裂解液也不能合成 DNA,当加入正常细菌的 DNA 聚合酶 III 时可以恢复它的聚合能力,因此不同于 DNA 聚合酶 I 和 II,聚合酶 III 是体内 DNA 复制所必需的酶。虽然每个大肠杆菌只含 10 个 DNA 聚合酶 III 分子,然而它的合成速率达到了体内 DNA 合成的速率。类似 DNA 聚合酶 I,DNA 聚合酶 III 具有 3′→5 和 5′→3′ 的外切酶活力。

纯的 DNA 聚合酶 III 不能利用带有引物的长的单链 DNA 作为底物,而也象 DNA 聚合酶 II 一样,可利用带缺口的双链 DNA 作为底物,但是"天然的"聚合酶 III (即下面所说的 Pol III*) 能利用单链 DNA 作为模板。

大肠杆菌 DNA 聚合酶 III 全酶由多种亚基组成,这些亚基很容易解离,而且在每个大肠杆菌中只有 10 到 20 个酶分子,因此不易获得均一的纯酶。仅在不久前才对其性质有一定了解。但每一组分起什么作用还不十分清楚。在 DNA Pol III 全酶中,由 α、ε、θ 三种亚基组成全酶的核心部分 (Pol III),但此核心只有很低的酶促活性,加上 τ 亚基后,组成的 Pol III′ 就可以利用长的单链 DNA 为模板了。但 Pol III 在 SSB 包围了单链 DNA 模板后就完全不能表现出酶活性来,因此各个亚基都有一定用处。Pol III,Pol III′ 和 Pol III* 都是不完整的 Pol III 全酶(见表 4-3,参阅表 4-1)(第 105 页)。 Pol III 全酶由 1 个 α 亚基、1 个 ε、2 个 θ、1 个 τ、2 个 γ、1 个 δ 和 1 个 β 共九个亚基组成。

在复制时 DNA 聚合酶 III 全酶十分重要,它可以在 DNA 的单链模板上,在引物的 3′OH 上以每分钟约 60,000 个核苷酸的速率逐个延长新生的 DNA 链。三种酶的性质比较见表 4-4。

表 4-3 DNA 聚合酶 III 全酶的组成

亚　基	分子量×10³	其 他 名 称			
α	140	dnaE 蛋白，pol C蛋白	Pol III	Pol III'	Pol III*
ε	25				
θ	10				
τ	83				
γ	52	dnaZ 蛋白	因子 II		
δ	32	因子 III，dna X 蛋白			
β	37	因子 I,dnaN 蛋白，copol III*			

表 4-4 三种大肠杆菌 DNA 聚合酶性质的比较

	Pol I	Pol II	Pol III
功能：5'→3' 聚合	+	+	+
3'→5' 外切	+	+	+
5'→3' 外切	+	−	+
焦磷酸化和 PP_i 交换	+		+
模板-引物:			
完整双链	−	−	−
有引物的单链	+	−	−
有断口的双链(多聚 dAT)	+	−	−
有缺口的双链或 5' 端突出的单链:			
单链长度<100 核苷酸	+	+	+
单链长度>100 核苷酸	+	−	−
分子量×1000	109	120	140
每个菌体中的分子数(估计值)	400		10—20
结构基因	pol A	pol B	pol C

真核生物中的 DNA 聚合酶

所有真核生物中的聚合酶的性质基本上同大肠杆菌中的聚合酶相似，它的活性有赖于模板,带 3'-OH 的引物,四种脱氧核糖核苷三磷酸。动物细胞的 DNA 聚合酶活性分布在好几个细胞级分中，但至今只对可溶性的 DNA 聚合酶研究得比较多。DNA 聚合酶 α 在迅速增殖的细胞中活力最高。这个酶的分子量在 165,000—175,000 之间。它的最合适的底物是带缺口的双链 DNA，但是带引物的变性 DNA 也是很好的底物。它同大肠杆菌 DNA 聚合酶最大不同之处在于不含外切酶活性。

DNA 聚合酶 β 是一个小分子量(43,000)的聚合酶。在整个细胞周期中它的活力没有变化,它利用 DNase I 处理过的天然 DNA 作模板,对变性的 DNA 几乎没有活力,这个酶也没有外切酶的活力。

DNA 聚合酶 γ 是一个大分子酶,在细胞内含量很低，因此不容易纯化。这个酶可以使用带缺口的 DNA 作为底物,但用多聚 (A)：寡聚 $(dT)_{8-10}$ 作为底物，酶活力更高，因此象是一个反向转录酶。

DNA 聚合酶 α 可能是真核生物的复制酶,主要的根据是聚合酶 α 的活性随细胞周期

而变化，如在淋巴细胞中，聚合酶 α 的活力和 DNA 合成的周期是一致的。而聚合酶 β 的活力升高要滞后一段时间，在 HeLa 细胞中，聚合酶 α 的活力从G期开始升高10倍，一旦S期结束，酶活力也降低，而聚合酶 β 在整个过程中没有变化。

DNA 连接酶

DNA 连接酶催化一个 DNA 链的5′磷酸根与另一个 DNA 链的3′羟基形成磷酸二酯键，但是这二个链必需都与同一个互补链结合，二个链必须是相邻近的。连接所需的能量或来自 NAD^+（如大肠杆菌)或ATP（如 T4 诱导的连接酶和动物细胞中的连接酶)。T4 诱导的连接酶不仅能在模板上连接 DNA 和 DNA 之间的切口，也能连接 RNA—RNA，RNA—DNA、和 DNA—RNA 之间的切口。大肠杆菌连接酶只能连接 DNA 和 RNA—DNA 之间的切口。T4DNA 连接酶不仅能连接双链 DNA 中的单链切口或双链的粘性末端，而且能连接无单链粘性末端的平头双链 DNA，大肠杆菌连接酶则不行。酶的反应机制是 ATP(或 NAD^+) 先与酶连接成共价中间体 E-AMP。E-AMP 和产物形成高能的焦磷酸键 Ap-p-DNA，这个中间体和相邻的 DNA 的 3′-OH 结合成磷酸二酯键(见第十章)。

DNA 连接酶在 DNA 复制、重组、修复中的作用是明显的，缺陷 DNA 连接酶的大肠杆菌突变株中冈崎片段积累，对紫外光敏感性增加，每个大肠杆菌约有300个连接酶分子，在30℃每分钟能连接7,500个切口，而在实际过程中，每分钟大约只有200切口需要连接，因此连接酶的含量对 DNA 复制而言是足够的。

三、DNA 双螺旋的解开

DNA 是双螺旋，在复制中双螺旋要解开，而螺旋的解开使得复制叉前面产生正超螺旋，如果这些应力不以某种方式释放将妨碍复制叉前进；如果让这些超螺旋传到分子末端才释放，对大肠杆菌而言，DNA 分子必须每分钟要转20,000转左右。而且环状分子还不能以这个方式释放正的超螺旋，除非分子内有一个切口使链可以自由转动。现在已找到一些酶和蛋白质，它们或者能使 DNA 双链变得易于解开，或者可以使超螺旋分子松弛，以下将介绍一些这类蛋白。

单链结合蛋白 (single-strand binding protein，简称 SSB)

最初发现噬菌体 T4 的基因32蛋白可以在远低于 T_m 的温度下使双链 DNA 拆开，并牢牢地结合在单链 DNA 上，后来发现多种生物中都有这类蛋白。在文献中，这类蛋白曾使用过多种名称：解旋蛋白 (unwinding protein)、熔化蛋白 (melting protein)，螺旋降稳蛋白 (helix destabilizing protein)。

从原核生物得到的 SSB 与 DNA 结合时表现出协同效应（图4-19）。比如第一个 SSB 结合到 DNA 上去的能力为1，第二个 SSB 结合能力高达 10^3，这可能由于蛋白质和蛋白质之间的互相作用增强了结合能力，也可能由于第一个 SSB 的结合使 DNA 链的结构发生了变化，有利于后来的 SSB 与 DNA 相结合，极有可能二者兼而有之。

从真核生物如牛、鼠和人细胞中得到的 SSB 与单链 DNA 结合时没有表现出协同

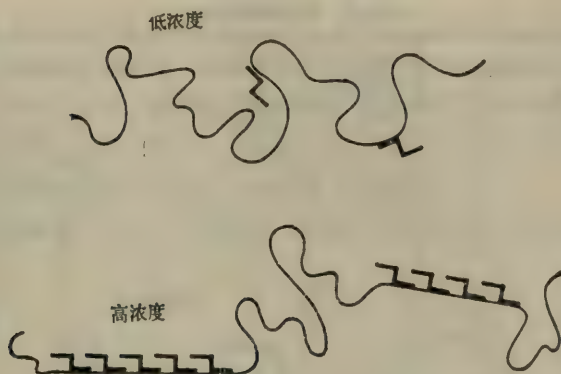


图 4-19 原核生物单链结合蛋白的协同作用

效应。

SSB 能刺激同源的 DNA 聚合酶活力。如 T4 的基因 32 蛋白能刺激 T4DNA 聚合酶的活力，但不能刺激大肠杆菌 DNA 聚合酶的活力，反之亦然。

SSB 能使天然 DNA 的融点降低，这可能不是它在体内的唯一生物功能。单链 DNA 与 SSB 结合之后能抵抗核酸水解酶的作用。如大肠杆菌 SSB 能保护单链 DNA 不受脱氧核糖核酸酶、蛇毒磷酸二酯酶、S_1 核酸酶、红色链胞霉内切酶以及 DNA 聚合酶 I 和 III 的 3′→5′ 外切酶的降解，但对小球菌核酸酶、大肠杆菌外切酶 I 仍然敏感，因此 SSB 的另一个重要生物功能可能就在于保护复制中的 DNA 单链部分不被核酸酶降解。

真核生物的 SSB 和原核生物的 SSB 可能有些不同，原核的 SSB 同双链 DNA 结合的能力很弱，而真核的 SSB 有一定结合能力；原核的 SSB 能促使变性 DNA 复性，而真核的 SSB 没有这种能力。因此原核 SSB 的作用方式是它先结合已存在的单链区，并从此侵蚀双链部分，而真核的 SSB 是先结合 DNA 双链并使之融化(图 4-20)。

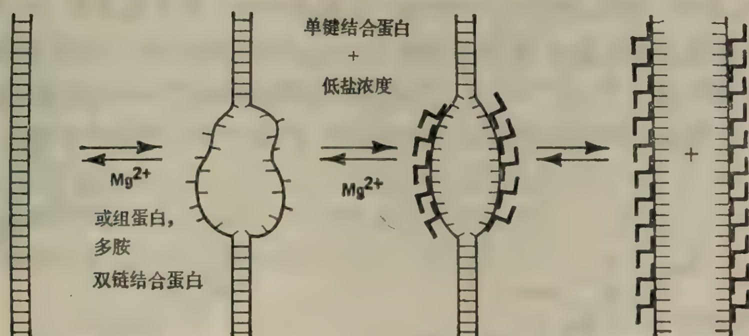


图 4-20 单链结合蛋白的侵蚀作用

DNA 解链酶 (DNA helicase)

有一类能通过水解 ATP 获得能量来解开双链 DNA 的酶称为 DNA 解链酶。表 4-5 列出了这些酶的部分性质。

解链酶分解 ATP 的活力依赖于单链 DNA 的存在。如果双链 DNA 中有单链末

表 4-5　解链酶的性质

解链酶来源	分子量×1000	解双链的能力	方　向	酶分子数/细胞	功　能
大肠杆菌 I	180	+	5′→3′	600	未知
大肠杆菌 II	75	−	5′→3′	6000	复制（大肠杆菌、λ、Col E1）
大肠杆菌 III	20	+	5′→3′	20	未知
大肠杆菌 rep	65	+	3′→5′	50	复制（φX174，M13）
大肠杆菌引发体(n′, dna B)	>500	+	5′→3′	20	引发
大肠杆菌 rec BC	268				重组
大肠杆菌 rec A	38			2000	重组
T4 基因 41	58		5′→3′		引发
T4 dda	56	−		6000	未知
T7 基因 4	57.66	+	5′→3′		引发
人细胞	110		5′→3′		未知

端或缺口，则 DNA 解链酶可以首先结合在这一部分，然后逐步向双链方向移动。复制时大部分 DNA 解链酶可以沿着后随链模板的 5′→3′ 方向随着复制叉的前进而移动，只有 rep 蛋白是沿前导链模板的 3′→5′ 方向移动。因此在复制时很可能 rep 蛋白和某种

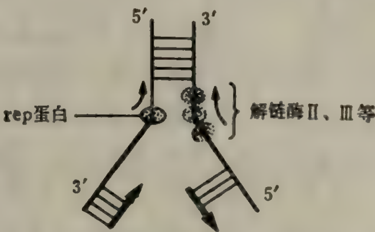


图 4-21　解链蛋白和 rep 蛋白的协同作用

DNA 解链酶分别在 DNA 的两条母链上协同作用，以解开双链 DNA（图 4-21）。

已经发现，有些 DNA 解链酶同时还有引发酶的功能，例如噬菌体 T7 的基因 4 蛋白可以优先与单链 DNA 结合，并在母链的 5′→3′ 方向靠水解 NTP 而移动并解开双链，直到 DNA 母链的某一专一识别位置，这时它就可以按模板的要求而合成小片段的RNA作为后随链的引物，比较多的引物链是 pppACCC 或 pppACCA（图 4-22）。

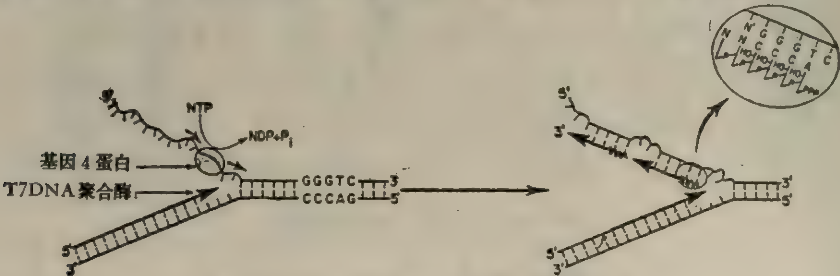


图 4-22　T7 基因 4 蛋白的解链作用和引发作用

拓扑异构酶 I（type I topoisomerase，简称 Topo I）

最初从大肠杆菌中分离到一种酶叫 ω 蛋白，它可以将共价闭环状的超螺旋 DNA 变

为松弛态，而并不改变其化学组成或结构。由于反应的产物还是闭环状 DNA，因此酶反应的机制只能是先切开双链 DNA 中的一条链，使链的末端沿螺旋轴按拧松超螺旋的方向转动，然后再把切口封起来。后来发现这种酶广泛存在于各种生物中，不同学者曾分别给以各种命名：转轴酶（swivelase），松旋酶（untwisting enzyme），DNA 松弛酶（DNA relaxing enzyme），切割封口酶（nicking-Closing enzyme）。

Topo I 酶松弛超螺旋 DNA 时不需要 ATP，它可以催化三种拓扑转换反应（图 4-23）。

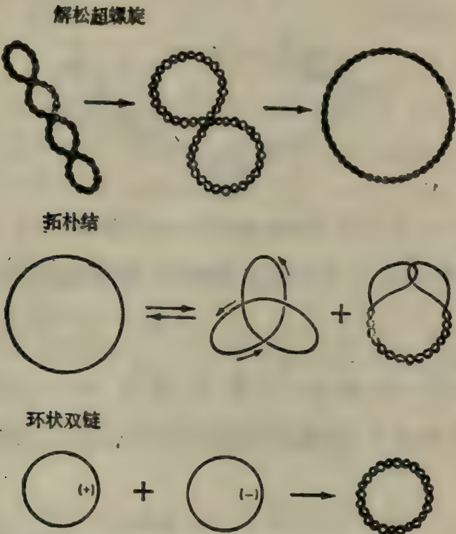


图 4-23　拓扑异构酶 I 的作用

Topo I 酶在复制中的作用还不十分清楚，某一大肠杆菌的突变株是没有 Topo I 酶基因的，但它的生长速度只比正常细菌稍稍慢一点，因此它并不是细菌生长必不可少的酶。

拓扑异构酶 II（type II topoisomerase，简称 Topo II）

拓扑异构酶 II 又称 DNA 旋转酶（DNA gyrase）或紧旋拓扑异构酶（twisting topoisomerase）。它由两个 α 亚基和两个 β 亚基组成。

拓扑异构酶 II 可以在水解 ATP 的同时使松弛态环状 DNA 转变为负超螺旋 DNA，其反应机制不同于拓扑异构酶 I，每次引起的超螺旋数变化为 2（见图 4-24）。这种反应说明拓扑异构酶 II 每次能使 DNA 的两条链同时切断，让另一条双链 DNA 穿过断口，再将已切断的末端重新封接起来，其反应过程可参考图 4-25。

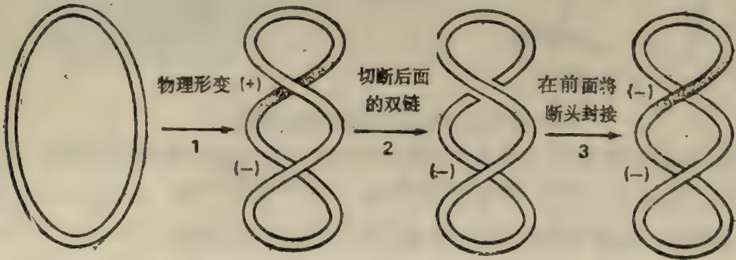


图 4-24　拓扑异构酶 II 的符号变换反应

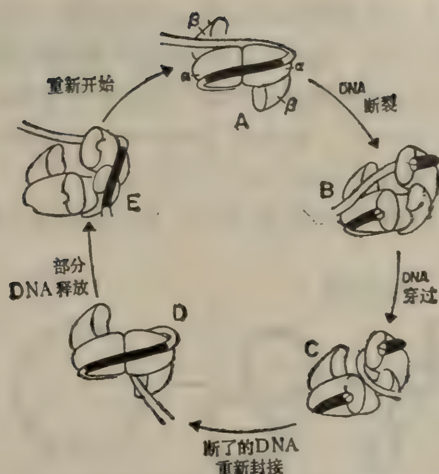
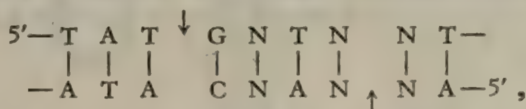


图 4-25 拓扑异构酶 II 的作用机制

拓扑异构酶不能非常专一地识别 DNA 的顺序，但常见的结构是



在加工过程中，两条链也不是平头的(临时切断处用箭头表示)，两条链之间相差四个核苷酸。

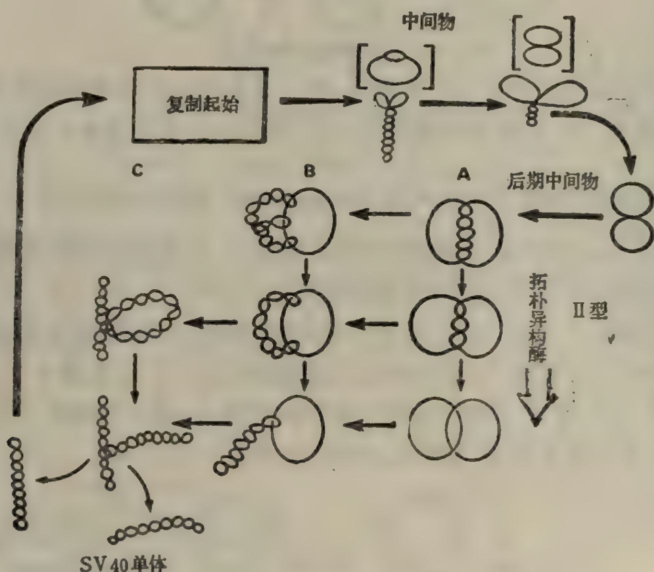


图 4-26 Topo II 的功能示意图

垂直箭号表示在 Topo II 作用下，逐渐拓扑异构化过程；横向箭号表示 DNA 合成；最后得到二个分开的超螺旋单体。

Topo II 在没有 ATP 时，可使超螺旋 DNA 变为松弛态。

Topo II 广泛存在于各种生物中，其功能十分重要，它是完成复制所必须的一种酶，

至少可以使完成了复制的两个子代环状双链 DNA 互相分开。图 4-26 表示 SV40 DNA 在 Topo II 作用下的复制与子代 DNA 的分开。

四、真核生物 DNA 的复制

真核生物的 DNA 比原核的大得多，而且都和组蛋白一起构成核小体。因此真核生物 DNA 的复制有一些原核生物所没有的特点。

核小体和染色体

真核生物的 DNA 通常总与许多蛋白结合在一起，以染色质的形式存在于细胞核中，在染色质中组蛋白的重量约占一半左右。组蛋白的组成和结构都是非常保守的，几乎所有的真核生物中都可以找到五种组蛋白：即 HI，H2A，H2B，H3 和 H4，它们在不同的生物和不同的器官中都有十分相似的氨基酸排列顺序。其中 H4 特别保守，牛、鼠和猪的 H4 性质完全相同，其结构与豌豆的 H4 相比也只有两个氨基酸不同：缬氨酸为异亮氨酸所取代，赖氨酸为精氨酸所取代。

染色质是由核小体串连而成的。在第二章已经介绍过，核小体结构的发现是近年来生物化学中很重要的进展之一，A. Klug 博士为此获得了 1982 年的诺贝尔奖。

染色质中 DNA 接头区是十分柔软的，因此可以允许串珠状的染色质按某种方式进一步绕。

细胞核内的染色质非常细，只有在电子显微镜下才能看到。但是到了细胞分裂期，在核内的染色质会发生形态和结构的重大变化，形成高度紧密的染色体（图 4-27）。经碱性染料染色后，即使在光学显微镜下也可以看到染色体的棒状小体。

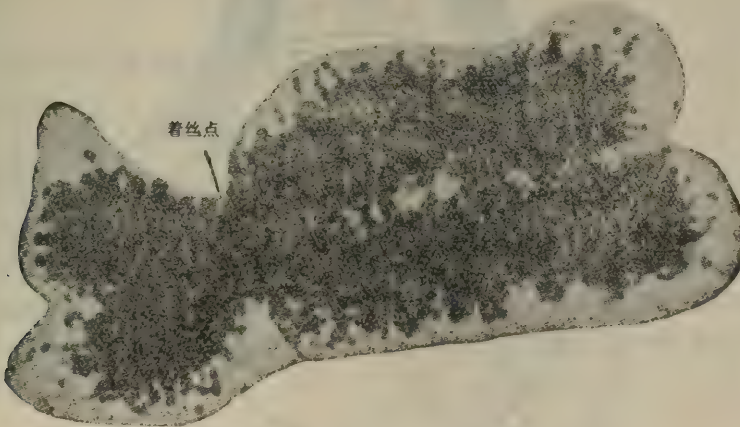


图 4-27 人染色体 12 的电子显微镜照片

染色体以两根平行的染色丝（chromonema）为主体，其中有一个不易染色的小颗粒叫着丝点（centromere）。整个染色体由一条连续的双链 DNA 贯穿着，即使在着丝点也不断。动物细胞每条染色体中 DNA 的长度不一，长的可达 4 厘米左右，其分子量约为 80×10^9。

各种生物的细胞核中染色体的数目是不同的。人的细胞核中共有 46 条染色体，其中两条是性染色体，人的染色体 DNA 长度约为 1.6 厘米到 8.2 厘米（相当于分子量 $3.2 \times 10^9 - 160 \times 10^9$）。

细胞周期和复制起始

真核细胞的生命周期可以分为四个时期：G_1、S、G_2 和 M，如图 4-28 所示。这四个时期所经历的时间明显不同，动物细胞组织培养时 S、G_2 和M期的三段时间是相对稳定的。S 期即 DNA 复制的时间约为 7 小时，G_2 期即有丝分裂前的准备期约为 4 小时，有丝分裂过程为 1 小时，而 G_1 期所需要的时间较长并有很大的变动范围，一般在 12 小时以上，但也有短的，例如酵母菌可以在 1.4 小时内繁殖一代。

在分化的细胞中，G_1 期可以进行 DNA 复制前准备，而在非分化的细胞中，可以存在一个性质上完全不同的 G_0 期，这时细胞处于静止状态，成年组织大部分细胞都处于 G_0 期。

从 G_0 转为 G_1 的关键反应是什么目前还不清楚，但有些化合物如 Ap_4A（见第九章第五节）可以明显地促进细胞由 G_0 期转入 G_1 期。G_1 期对细胞的生长周期是非常重要的。

真核生物染色质的复制与原核生物 DNA 的复制有三个明显的不同之处：(1)真核生物每条染色质上可以有多处起始点，而原核生物总是只有一个起始点；(2)真核生物的染色体在全部复制完成之前，各个起始点上 DNA 的复制不能再开始，而在快速生长的

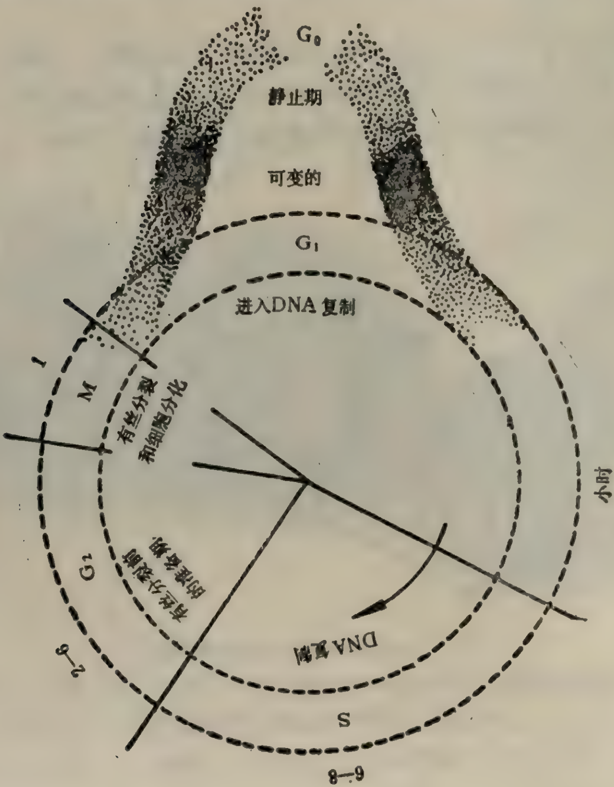


图 4-28 真核细胞的生长周期

原核生物中，在起始点上可以连续地开始新的 DNA 复制；(3)真核生物 DNA 复制起始点并不固定在某一 DNA 顺序处，在酵母基因组的 400 多处有一种自发复制顺序 ars(autonomous replicating sequence)，这种顺序广泛存在于线粒体中，但在大肠杆菌中却是找不到的。

DNA 复制单元或复制子（DNA replication unit 或 replicon）

如上所述，原核生物如大肠杆菌DNA 复制时只有一个复制起始点，然而在真核 DNA 的复制时有多个复制起始点。在哺乳动物复制中的 DNA 放射自显影图上可以看到很多复制眼，每个复制眼都有独立的起点。双向延长（见图 4-6 III)。每一个这样的 DNA 单位称为复制单元。复制单元大小是不均一的，大约 15—300μm(45—900kb)，大多数在 10—100μm 之间。各种不同的生物其复制单元的大小有区别，就是同一种生物在不同的生长条件下复制单元的大小也有区别，同一个 DNA 上不同的复制单元大小也不均一，生长比较快的细胞，复制单元也短一些。另外病毒感染也影响复制单元大小，例如 SV40 转化的中国田鼠细胞的起始间隔就比未转化的要小。如果用 SV40 的T抗原温度突变株转化的细胞，在允许温度复制单元要短些，而在限制温度复制单元的大小恢复到正常，这些实验都指出，复制单元的大小受功能方面的因素影响。

另一方面，有些工作者认为复制单元的大小可能受染色体的结构控制，在果蝇的多线染色体中，一个横带所含的 DNA 就是一个复制单元。

几个邻近复制单元可能组成复制单元簇，复制单元簇可以由少至二个复制单元多达 250 个复制单元构成。复制单元簇是受协同调节的，不同的复制单元簇在开始复制的时间上有先后，同一复制单元簇各个复制单元基本是同步的，若用射线或抑制剂抑制 DNA 合成的起始，则同一复制单元簇呈协同一致的反应。

染色体的合成

在真核生物 DNA 复制的同时组蛋白也在不断产生，因此新生成的 DNA 可以很快地与组蛋白结合在一起。

在染色体复制过程中亲代组蛋白是同新合成的组蛋白打乱重新组合分配到子代中，还是亲代组蛋白作为一个整体传下去？用鸡肌母细胞作为材料，用[^{3}H]赖氨酸和 ^{13}C 及 ^{15}N 氨基酸标记染色体，从这种细胞中分离核小体，每个核小体中的组蛋白用化学方法交联，去除 DNA 后分析交联的组蛋白密度，尽管新合成的组蛋白仅占总组蛋白七分之一，然而所有新合成的组蛋白都含有标记同位素，这表明新合成的组蛋白和老的组蛋白是不混杂的。这个实验证明核小体复制进程中，DNA 复制是半保留的，而组蛋白却以全保留的方式传给子代。

五、DNA 复制的调控

分子生物学家们并不满足于对 DNA 复制过程的观察和描述，而是希望了解复制过程中各种酶和蛋白质的作用，它们与 DNA 的关系以及复制是怎样受到调节和控制的 。

现以 λ DNA 为例来说明原核生物中 DNA 复制是怎样调节的。

噬菌体 λ 以及其他若干种温和噬菌体是非常令人感兴趣的，它们的 DNA 不但可以自主地进行复制，也可以转变为溶原状态，即插入到大肠杆菌的染色体中，从而与寄主处于共生的潜伏状态，这时 λ DNA 的基因都不积极活动（见第二章）。因此这类温和噬菌体成为我们了解原核生物中 DNA 复制调控过程的良好实验材料。

噬菌体 λ 是一个带有长尾的 20 面体，头部蛋白外壳中含 48Kb 长的一条 DNA。λ DNA 线状分子两端有 12 个核苷酸长度的粘性末端(图 4-29)。

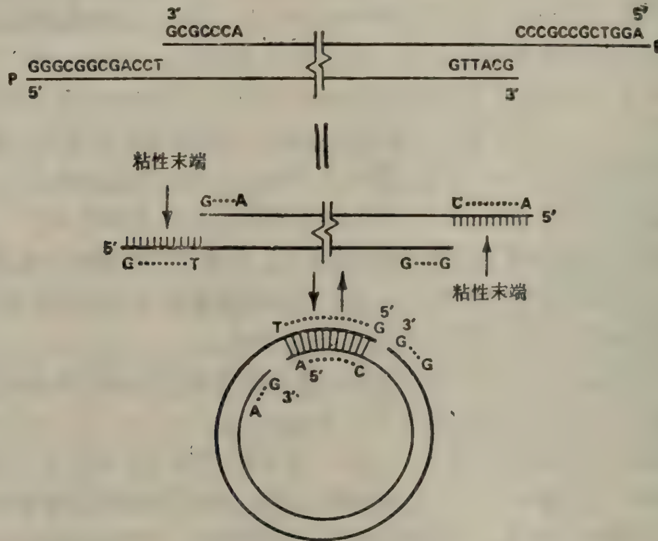


图 4-29　λ DNA 的粘性末端

噬菌体 λ 遇到大肠杆菌后，其尾部的三个蛋白分子（为 λ DNA 中 J 基因所编码）可以专一地识别寄主的菌壁。噬菌体 λ 将其 DNA 注入大肠杆菌后，线状的 λ DNA 分子立刻靠粘性末端、连接酶、拓扑异构酶 II 以某种尚不知道的机制很快地转变为超螺旋状态。这时寄主就面临着两种完全不同的命运。一种是 λ DNA 进行烈性繁殖，产生上百个有感染活力的新病毒，直接导致寄主的裂解和死亡，另一种是 λ DNA 插入寄主的“染色体” DNA 中。两条道路的选择取决于当时寄主和噬菌体中的多种因素，这些因素的微妙平衡综合在一起可以决定基因采用那种表达方式。

Jacob 等人首先研究了复制子的调制。所谓复制子是遗传上的一个复制单元，其中有一个特定的复制起始点可以作为一个整体而被复制。最简单的复制子之一是噬菌体 λ 中占 10% 的一段 DNA，称为 λdv。λdv 是人工制造的环状 DNA （见图 4-30）。它包括 O_L、P_L、cI、O_R、P_R、cro、cII、ori、O、P 等基因，其中 cII 不是自我复制所必需的基因，但 P_R、O_R、cro、O、P、ori 基因都是必不可少的。它与转录的控制基因紧挨在一起（见第二章图 2-11）。这是一个能自我调控的复制子，这个调控是通过基因 cro 的产物 cro 蛋白结合在 P_R、O_R 处，能阻止 cro 蛋白本身的产生，同时也能调节基因 P、O 二个产物的生成，因为这二个基因也是从 P_RO_R 开始转录的，转录本身也是复制起始所必须的，起始点附近没有这转录起始点就不能活化，有关的机制还不清楚。基因 O 和 P 的产物是复制的引发蛋白，因为基因 cro 同基因 O 和 P 是同步转录的。从 P_R 转录一次，不仅产生新的 O 和 P 蛋白

使 DNA 复制一次，而且产生更多的 cro 蛋白，而 cro 蛋白浓度增加，作为一个阻遏蛋白阻止了新的 O 和 P 蛋白的产生。因此 cro 又是这个复制子的自我调节基因。用 cro 的温度敏感株做实验，表明 λdv 在细胞内的数目同 cro 的浓度直接有关，在限制温度培养时 DNA 的合成速率大大增加，这种情况对宿主细菌是致死的，可能宿主容纳不了这么多质粒。这个结果指出，cro 对 λdv 在细菌内的维持是必要的。

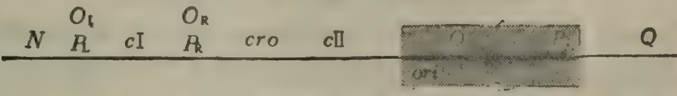


图 4-30 λ DNA 的复制控制区

λ DNA 带有较多的基因，因此复制的调控比较复杂。现在已经明确，λ DNA 进入大肠杆菌后不久，菌体内就会出现若干种阻遏蛋白。它们是 cI 基因和 cro 基因的产物，能作用在 λ DNA 的 O_R 和 O_L 位点上阻止 λ DNA 的转录(参见第二章)，阻遏蛋白

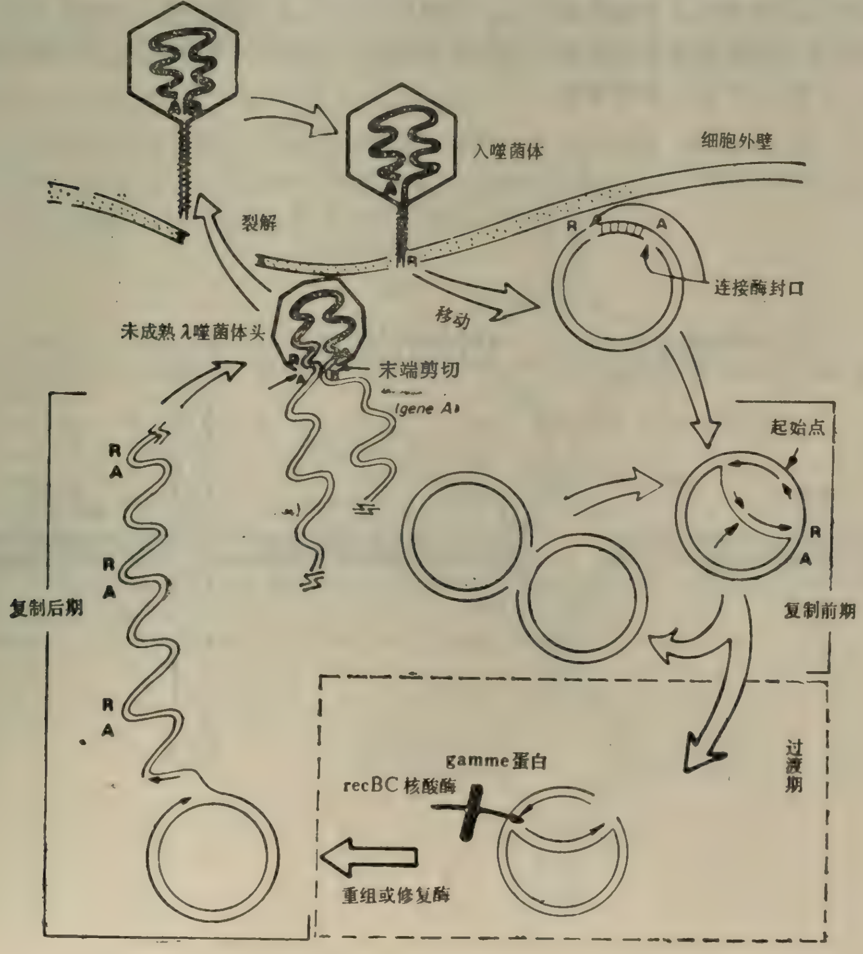


图 4-31 噬菌体 λ 复制的方式前后期不同

的多少可以通过调节O蛋白的含量来决定λDNA 采用那一种复制方式。在寄主受到噬菌体λ感传的早期O蛋白浓度较高,这时 λDNA 可进行双向复制。在后期O蛋白浓度较低,因而 λDNA 以单向的滚环方式复制(图 4-31 是总的示意图)。早期 λDNA 的复制以θ方式进行。这时复制需要O基因和P基因的产物,以及寄主的蛋白如 RNA 聚合酶, dna B 蛋白,引发酶和 DNA 聚合酶 III 等。近年来O蛋白和P蛋白已分别用遗传工程技术提纯,分子量分别为 36×10^3 和 26×10^3。纯化的O蛋白和P蛋白加到 ori C 酶系统后,可以从一个λDNA 的起始区上引起很激烈的复制。

在复制后期, gam 基因产生的γ蛋白可以抑制 rec BC 基因编码的核酸酶含量,因而可以使λDNA 按滚环方式复制 DNA。然后在基因 Nu I 产物和寄主中其他因子的帮助下,成熟的多聚λDNA 可以和 5 种外壳蛋白(分别由基因 B. C. N3. D. E 编码)组装成噬菌体λ的头部。每个头部只能容纳一条 λDNA 单体,多余的部分由基因A的产物加工,切出带粘性末端的单体,最后形成完整的噬菌体λ。

当 int 基因产生足够的 int 蛋白时,超螺旋 λDNA 才会整合到寄主 DNA 中去。一旦λDNA 进入溶原状态,低水平的 cI 阻遏蛋白就足以阻止溶菌的发生,并保持 cI 阻遏蛋白的合成维持在比较低的水平上。在溶原状态时,λ基因与寄主 DNA 一同复制。只有在受到外来因子如紫外线照射、胸腺嘧啶供应不足或丝裂霉素破坏了阻遏蛋白的抑制状态时,λDNA 才会大量地增生。

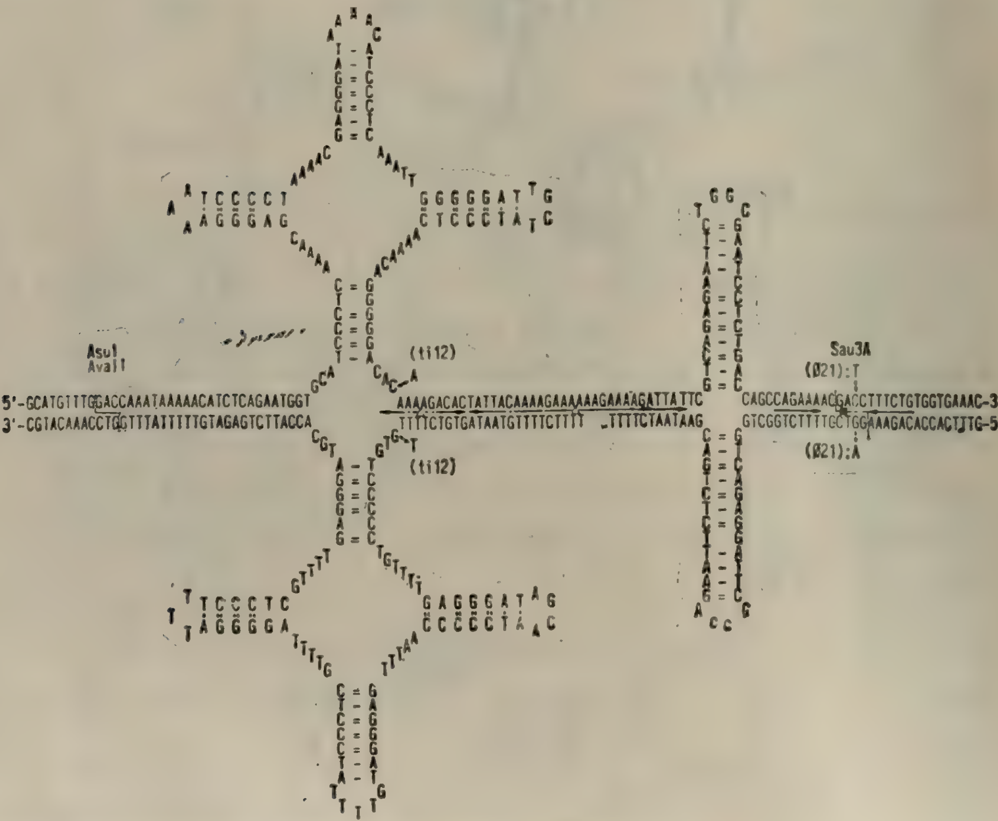


图 4-32 噬菌体λ复制起始点的顺序和二级结构

一旦 λ DNA 进行恶性繁殖,在早期阶段超螺旋 λ DNA 从O基因中的特定起始点上开始从两个方向复制 DNA。早期复制所需的一切必要基因都在起始点附近 2,500 碱基对范围以内,其结构如图 4-32 所示,呈四叶草形,并能为某种蛋白专一地识别。

通过对噬菌体和病毒的研究,人们现在对原核生物中 DNA 的调控过程有了一些了解,当然原核生物 DNA 复制要比病毒和噬菌体复杂得多。

原核生物的生长和繁殖可以有不同的速率,这取决于培养的条件,但是一旦 DNA 开始复制, DNA 链的延伸速率却是相对恒定的。例如大肠杆菌一般在 37℃ 时从 DNA 复制起始到完成一般要 40 分钟,在两个基因组之间形成膜并分裂为两个子细胞要 20 分钟,两者几乎是恒定不变的。细菌生长周期中变动最大的是复制的起始,新的 DNA 可以在前一条链尚未完成之前就又开始复制的起始,也可以在细胞分裂以后很久才发生,这与细菌中各种激发因子、抑制因子、去抑制因子的复杂相互影响有关,染色体复制同细菌的分裂也直接有关,用氯霉素抑制蛋白质合成,染色体复制不完全,最后有 0.5% 是不复制的,而这些染色体复制的完成是细胞分裂的信号。

真核生物 DNA 的复制又比原核生物复杂得多,至今我们还不清楚真核生物中 DNA 复制是怎样调控的,为什么胚胎细胞会迅速地生长而成年细胞生长很缓慢,有的成年细胞甚至在几十年里都不生长,什么因素会使一个正常细胞转变为恶性生长的肿瘤细胞,这些都是生命世界向人类提出的带有挑战性的问题。

近年来真核生物 DNA 的复制和基因的调控正受到越来越大的重视,很可能这方面的研究会为细胞分化、再生、衰老等现象提供生物化学的说明,也可能为肿瘤的预防和治疗提供有价值的线索,从而为生物世纪的来临铺平道路。

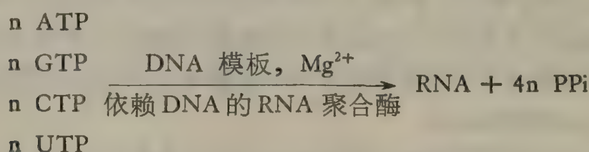
参　考　文　献

[1] Kornbery, A. DNA Replication, W. H. Frecman and Company, San Francisco, 1980.
[2] Kornbery, A. Supplement to DNA Replication, W. H. Frecman and Company, San Francisco, 1982.
[3] Smith-keary, P. F. Genetic structure and function. Macmillan Press Ltd. 1975.
[4] Nossal, N. G. 1983, *Ann. Rev. Biochem.*, **52**, 581.
[5] Challberg, M. D. and Kelly, T. J. 1982, *Ann. Rev. Biochem.*, **51**, 901.
[6] Itobom, G. 1981, *Curr. Top. Microbiol. Immunol.*, **94**, **93**.
[7] McHenry, C. and Kornbery. A. 1981, *Enzymes*, **14**, 39
[8] Kolter, R. and Helinski, D. R. 1979, *Ann. Rev. Genet.*, **13**, 355.
[9] Wood, W. B. et al., 1981, Biochemistry, A Problems Approach, 2nd ed., The Benjamin Cummigs Publishing Company.
[10] Crow, J. F. 1983, Genetics Notes, 8th ed. Burgess Publishing Company.

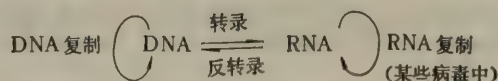
第五章　转录与加工

祁国荣　谌章群　辜祥荣

核酸和蛋白质是组成生物体的两大类重要物质，它们相互依存，彼此制约。1958 年 Crick 提出的中心法则，提出信息流向为 DNA→RNA→ 蛋白质，揭示了核酸与蛋白质的内在关系，核酸(主要是 DNA)是信息的贮存者，而蛋白质则是信息的展现者，由于蛋白质的活动，表现出生物体的各种性状和功能。遗传信息从 DNA 经过 RNA 传递到蛋白质，称为基因表达，基因表达的第一步，即信息从 DNA 到 RNA 的转移，叫做转录。转录过程是以 DNA 为模板，在依赖 DNA 的 RNA 聚合酶催化下，由四种核糖核苷三磷酸 (NTP) 合成 RNA 的过程：



转录是依赖 DNA 的或叫 DNA 指导的 RNA 生物合成，它不同于 RNA 病毒复制时依赖 RNA 的 RNA 生物合成，催化后者的酶叫依赖 RNA 的 RNA 聚合酶。某些 RNA 病毒(例如肿瘤病毒)可把信息从 RNA 流向 DNA，即依赖 RNA 的 DNA 生物合成，鉴于它与"转录"的对应关系，被叫做"反转录"，它们的关系可图示为：



关于 RNA 复制和反转录将在第八章讨论。

近一、二十年来在分子生物学上有一个重大的发现，即转录的原始产物是没有功能的 RNA 分子，它必须经过加工 (processing) 才能得到成熟产物，(原核生物 mRNA 的原始转录产物一般不需要加工是一个例外)。原始转录产物叫 RNA 前体，加工过程中还可能产生各种中间前体，最后才得到有功能的 RNA 分子。因此将加工和转录一起讨论是必要的。和转录有关的加工叫转录后加工 (post-transcriptional processing) 以区别于和翻译有关的加工——翻译后加工 (post-translational processing)(见第六章)。不同种类的 RNA 前体的加工是很不一样的。

一、RNA 聚合酶(依赖 DNA 的 RNA 聚合酶，转录酶)

转录是由 RNA 聚合酶来完成的。1955 年 Ochoa 等发现了多核苷酸磷酸化酶(PNPase)以后，曾一度认为 RNA 生物合成问题可以解决了，但是在仔细分析该酶的作用特点及所合成产物的特性以后，许多研究者对此发生了怀疑。Weiss 于 1959 年首先在鼠肝细胞核

系统中观察到 NTP 可参人到 RNA 产物中;其后不久,数家实验室从多种材料报道了依赖于 DNA 模板的 RNA 聚合酶活性,并进行了酶的初步提纯,发现合成的 RNA 产物与所加入的 DNA 模板是互补的。不同 DNA 来源对合成的 RNA 碱基组成的影响见表 5-1。

表 5-1 DNA 模板与 RNA 产物碱基对应关系

DNA 来源	DNA 中 $\frac{A+T}{C+G}$	RNA 中 $\frac{A+U}{C+G}$
T2 噬菌体	1.86	1.85
小牛胸腺	1.35	1.52
大肠杆菌	1.0	0.93
小球菌	0.40	0.48

大量研究表明，原核细胞与真核细胞的 RNA 聚合酶在结构上、性质上是很不相同的。

原核生物 RNA 聚合酶

迄今为止,大肠杆菌 RNA 聚合酶是研究得比较透彻的一个酶。大肠杆菌 RNA 聚合酶全酶 (holoenzyme) 可用磷酸纤维素柱层析分离为一个核心酶 (core enzyme) 和一个称为 σ 的多肽。核心酶又由 4 个亚基 ($\alpha_2\beta\beta'$) 组成,亚基及全酶的分子量见表 5-2。在用聚丙烯酰胺凝胶电泳测定分子量时，各家报道的分子量略有出入。此外还有一个低分子量的成份 ω(9,000—12,000d) 存在于某一些酶制剂中 (每个酶分子含 0.2— 2 个分子)。噬菌体 T3、T7 RNA 聚合酶只有一条肽链,分子量为 110 × 10^3d。

表 5-2 原核生物 RNA 聚合酶亚基组成

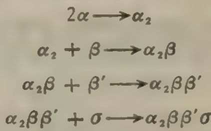
酶来源	全酶分子量 ($\times 10^{-3}$)	亚基结构	亚基分子量($\times 10^{-3}$)			
			α	β	β'	σ
大肠杆菌*	454	$\alpha_2\beta\beta'\sigma$	37	150	160	70
假单胞菌	500	$\alpha_2\beta\beta'\sigma$	44	155	165	98
枯草杆菌	420	$\alpha_2\beta\beta'\sigma$	45	155	155	57
Anacystis nidulans	436	$\alpha_2\beta\beta'\sigma$	39	147	125	86

* 大肠杆菌 RNA 聚合酶 α、β 和 σ 亚基的氨基酸顺序已经测出,分别为 329、1342 和 613 残基，准确分子量分别是 36512、150619 和 70263d。

在 RNA 聚合酶的亚基结构中，α、β 和 β' 亚基对酶活性的重组是必需的，ω 则是不必需的,其功能尚不清楚。核心酶具催化活性,而 σ 亚基本身没有催化活性，其作用是识别 DNA 分子上 RNA 合成的起始信号。但 σ 亚基不能单独与 DNA 结合,它结合到核心酶后可能引起酶构型的变化,因而改变核心酶与 DNA 结合的特性。根据单个亚基或单个亚基聚集体的功能特征、亚基的特异探针和变异体混合重组的研究，表明 β' 亚基是仅有的可以单独与 DNA 结合的亚基,它参与 RNA 聚合酶与模板反应,同时 β' 亚基也与核心酶和 σ 亚基结合以及转录的终止有关。β 亚基具有利福霉素和利链菌素 (见第三节)的结合点,也与 σ 和 α 亚基结合有关,它参与 RNA 链的引发和延伸,催化底物形成

磷酸二酯键。β(或β')也可与终止因子 ρ(见下面)发生直接反应。α 亚基具有与β的结合点,参与特定的基因表达,可能与酶和 DNA 上的启动区域的反应有关。

除了四个多肽亚基外,大肠杆菌 RNA 聚合酶还含2个 Zn 原子,它们与 β' 亚基相联接。无论在体外或体内,RNA 聚合酶通过下列四步反应聚集而成:



$$2\alpha \longrightarrow \alpha_2$$
$$\alpha_2 + \beta \longrightarrow \alpha_2\beta$$
$$\alpha_2\beta + \beta' \longrightarrow \alpha_2\beta\beta'$$
$$\alpha_2\beta\beta' + \sigma \longrightarrow \alpha_2\beta\beta'\sigma$$

细菌 RNA 聚合酶在体外低离子强度下有一种可逆的凝聚作用,例如大肠杆菌全酶在 0.1M 以上的离子强度时以单体存在,在较低的离子强度下则形成二聚体。核心酶在较低离子强度下则是以六聚以上的多聚体形式存在。所有 RNA 聚合酶都具有相同的功能——合成 RNA,但还可能与转录调节有关。如枯草杆菌芽孢化是与σ因子的失去和β' 亚基的修饰联系在一起的,因而枯草杆菌的 RNA 聚合酶突变株可阻止该菌的芽孢化。

原核生物 RNA 聚合酶能催化所有类别的 RNA 的生物合成,被利福平等抑制(见第三节)。

真核生物 RNA 聚合酶

与细菌仅含一种类型的 RNA 聚合酶不同,在动物、高等植物和真菌的细胞核中却有多种形式的 RNA 聚合酶。过去,这些 RNA 聚合酶常依其在细胞内的位置和模板特异性、柱层析行为、二价离子的活化作用、对特异性抑制剂 α-鹅膏蕈碱(α-amanitin)的敏感性以及亚基的组成等区分或命名,在文献中比较混乱,现统一的命名列于表 5-3。

表 5-3 真核生物 RNA 聚合酶

RNA 聚合酶	I(或 A)	II(或 B)	III(或 C)
别名	rRNA 聚合酶	不均一RNA 聚合酶	小分子 RNA 聚合酶
细胞内位置	核仁	核质	核质
转录产物(前体)	rRNA	mRNA	tRNA,5S rRNA
α-鹅膏蕈碱效应*	>10^{-3}M 抑制(不敏感)	10^{-9}—10^{-8}M 抑制(高度敏感)	10^{-5}—10^{-4}M 抑制(中度敏感)

* 昆虫和酵母的 RNA 聚合酶 III 对 α-鹅膏蕈碱是不敏感的,类似 RNA 聚合酶 I。

真核生物 RNA 聚合酶的亚基情况相当复杂,动植物来源的各举一例列于表 5-4。

从已经测定的真核生物 RNA 聚合酶的亚基组成分析,亚基数从 4—10 个不等,有的多至 14 个(SDS 凝胶电泳),种类有 4—6 种,其中有二种为大亚基(分子量>100×10^3),余为小亚基,可能出现小分子量的降解产物。真核 RNA 聚合酶也紧密结合 Zn 原子。从酵母和胸腺组织中可以分离出缺乏一种或二种小亚基的 RNA 聚合酶 A,仍具有酶活性,表明这些亚基对催化活性是不必要的,可能与转录调节有关,或功能上类似于原核生物的σ亚基,虽然尚无直接证据说明真核生物有类似于原核的σ亚基存在。实验指出,

RNA 聚合酶A和B中三个小亚基的胰蛋白酶水解的肽图是相似的，同时二者也具免疫相关性。

表5-4 牛胸腺和椰子 RNA 聚合酶的亚基和分子量

来　源	种　类	分子量 ($\times 10^{-3}$)	亚基数	亚基种类	比　例	亚基分子量 ($\times 10^{-3}$)
牛胸腺	A	501	8	A1	1	197
				A2	1	126
				A3	1	51
				A4	1	44
				A5	2	25
				A6	2	16.5
	B	538	10	B1	1	214
				B3	1	140
				B4	2	34
				B5	2	25
				B6	4	16.5
椰　子	A	630	6	a	1	180
				b	2	150
				c	1	69
				d	2	42
	B	730	6	a	1	180
				b	2	150
				c	1	95
				d	2	80
	C	534	6	a	1	180
				b	2	112
				c	1	50
				d	2	40

除了上述细胞核 RNA 聚合酶外，也分离出线粒体 RNA 聚合酶。叶绿体中也存在不同的酶。某些线粒体 RNA 聚合酶具有非常简单的结构，是由一种单一亚基聚集而成。线粒体和叶绿体 RNA 聚合酶不同于真核细胞核 RNA 聚合酶，而与原核生物相似，能催化所有类别 RNA 的生物合成。

二、转录的机制

关于转录机制的研究是从大肠杆菌开始并深入的，其它原核生物也有类似的特性。虽然哺乳动物细胞的转录调控和转录产物的加工与原核生物是不一样的，但这两类生物的转录过程却十分相象。

在 DNA 双链分子中，被转录为 RNA 的那一条链叫有意义链 (sense strand)，该基因上的另一条链则称反意义链 (antisense strand)(图 5-3)。在含有许多基因的 DNA 双链中，每个基因的有意义链并不总是在同一条 DNA 链上，就是说一条链上具有某些基因的有意义链和另一些基因的反意义链。这个现象的一些具体例子已在第二章介绍过。

转录可以分为起始、RNA 链的延伸和转录终止三个步骤。

转录的起始

虽然 RNA 聚合酶的核心酶在 DNA 模板、核苷三磷酸底物和 Mg^{2+} 存在下也可以合成 RNA，但产物是不均一的，没有固定的起始点，而且 DNA 双链的每一条链都能作为模板。RNA 聚合酶全酶则能专一地与 DNA 上 RNA 合成的起始位点（该位点被叫作启动基因或启动子）结合，全酶合成的 RNA 链是均一的，并且只选择一条 DNA 链作模板。那么，启动子究竟具有什么样的特殊结构使全酶能够识别呢？近年来，已测定出相当多的启动子的一级结构，这些结构通常是用下述方法分离得到的：在全酶与双链 DNA 结合后，用 DNase 水解，便可得到一个全酶保护的 DNA 片段。例如，fd、T_7A_2、T_7A_3、λ_{pr} 和 $LacUV_5$ DNA 的全酶保护片段有 41—44 个碱基对，在一定条件下，RNA 聚合酶全酶可以在被保护的 DNA 片段上引发 RNA 的合成，合成的产物链长约 17—20 核苷酸，并且对应于各自的 mRNA 5′ 端的起始顺序。1975 年 Pribnow 比较了上述五种 DNA 片段的核苷酸顺序，观察到有一个七个碱基的同源区，即 TATAATG，后被称为 Pribnow 盒，其中心是位于 mRNA 起始点前大约 10 个碱基对位置。以后，Schaller 等人发现，若把被全酶保护的 DNA 片段提纯，RNA 聚合酶全酶不能再重新结合或者不能再选择正确的引发点，对于这个现象的解释是，在被保护片段以外一定还有一些顺序对于酶识别启动子是必需的。从 λ 和 SV40 的研究表明，启动子区域的 DNA 顺序必需从 mRNA 的起始点向左延伸到 35 个碱基对以远，而乳糖操纵子的启动子区域必需从 RNA 合成起始点向左延伸到 60 个碱基对以远。在 λ 的左右启动子和 SV40 启动子的-35 位上有一个共同的顺序，它是 Hin 限制性内切酶的切点，用 Hin 酶切开后便失去了启动子活性。Schaller 和 Gilbert 比较了大肠杆菌 RNA 聚合酶的 46 个启动子结构，把 Pribnow 盒排在一起，发现它们都存在着两个同源区，即位于-10 碱基对左右的 TATAATG 顺序和位于-35 碱基对左右的 TTGACA 顺序(图 5-1)。这两个顺序中的碱基在每一个启动子中虽不尽相同，但它们具有很大的保守性。例如，在 Pribnow 盒中，所有启动子都具有其中第六位碱基 T (TATAATG)，TA 和 TAA 也有很强的保守性。-35 碱基对位置的同源顺序 TTGACA 中的 TTG 也是很保守的(图 5-1)。由此看来，启动子结构至少由三部分组成，第一部分是 RNA 聚合酶全酶的识别信号(-35 碱基对左右)；第二部分是酶的紧密结合点(-10 碱基对左右)；第三部分是 RNA 合成的引发点。一些化学和光化学探针对 RNA 聚合酶和启动子结合的研究揭示，两者之间的专一接触点，主要是上述两个同源区，但也存在于其他区域。有时正、负调节转录的蛋白结合位点也包括在启动子结构内(见第四节)。

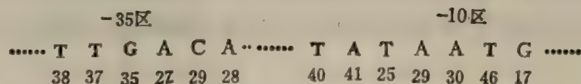


图 5-1 原核生物启动子中-10与-35 区域的同源性
比较了 46 个启动子，图中核苷酸下面数字表示出现的数目。参阅 *Ann. Rev. Genetics*, 1979, **13**,319—353。

关于原核生物 DNA 转录的起始过程，根据目前知识大致可概括如下：

（1）核心酶在 σ 亚基参与下与 DNA 分子接触，形成非专一复合物，这样的复合物是很不稳定的；

（2）起始识别，全酶与启动子结合生成封闭的启动子复合物（closed promoter complex），这时酶与 DNA 外部结合，识别部位大约在启动子的-35 位处；

（3）活化，得到开放的启动子复合物（open promoter complex），酶在-10 位处紧密与启动子结合，解开双链后识别有意义链。由于该部位富含 A-T 碱基对，故有利于 DNA 解链；

（4）形成三元起始复合物：酶移动至起始点，加入第一个核苷三磷酸得到(启动子·全酶·核苷三磷酸)复合物，即开始转录。

差不多所有情况下，RNA 链的开头碱基总是嘌呤，大多数大肠杆菌转录得到 RNA 的原始产物的第一个核苷酸是 pppG 而 ϕX174、T7DNA 为模板时，更多的起始核苷酸为 pppA。

以上是原核生物启动子的情况，近年来，对真核生物 DNA 上启动子的研究也发现，在转录起始点 5′ 端上游约 30 bp 处也存在一个 A-T 丰富的同源区-TATAAA(AG)，它与原核生物启动子-10bp 处的 Pribnow 盒很相似，被称作 Hogness 盒。真核 RNA 聚合酶 II 在此（T）ATA 盒处与 DNA 结合，然后在下游约 30bp 处转录 mRNA 前体。

比较了珠蛋白基因 5′端上游的核苷酸顺序观察到有如图 5-2 所示的规律性，从起始密码子 ATG 开始向左延伸，有一个同源 AC 顺序(此处转录生成帽子结构)，再左延伸出现一个 ATA 的保守同源区，为珠蛋白基因有效转录所必需，缺失 ATA，则转录起始点不准确，同时转录效率大大降低。此外，5′旁侧还有一个 CCAAT 同源顺序(见图 5-2)。

	CCAAT		ATA		帽子		起始密码
β	CCAAT	39	CATAAA	26	AC	50	ATG
δ	CCAAT	31	CATAAA	26	AC	50	ATG
α_2	CCAAT	34	CATAAA	23	AC	37	ATG

图 5-2 珠蛋白基因 5′ 端上游顺序的同源性

（数字代表核苷酸数目）

RNA 链的延伸

当转录起始生成第一个核苷酸，RNA 链延伸随之进行。从起始到延伸过程的转变，DNA 分子与酶分子可能发生构象变化，以适应延伸过程的需要。当酶从起始点往下游移动，解旋也随酶一起进行(图 5-3)。而原来部位则重新形成完整双螺旋，表明 DNA 螺旋在酶(核心酶)作用下有一个可逆过程，同时酶分子的构象也发生变化。在转录起始阶段，全酶与 DNA 形成稳定的复合物，对连接部位的一级结构有强的专一性；而在延伸阶段，为了能沿模板移动，它必须放松对 DNA 的结合，同时对拷贝每一个核苷酸应具同等能力，而不应有所偏向。RNA 聚合酶与 DNA 模板相互作用的这些变化，使得转录起始后立即从酶分子上释放 σ 亚基；σ 亚基存在与否，引起 β 与 β′ 亚基构象上的变化：即当 σ 亚基缔合时，β 与 β′ 表现为有利于专一与 DNA 结合的构象，而 σ 释放后留下核心酶，则与 DNA 结合不专一，失去识别或优先于任一核苷酸顺序的能力，不能形成稳定的复合物。转录起始至延伸的过程，包括一个 σ 因子的缔合与解离循环(见图 5-8)。

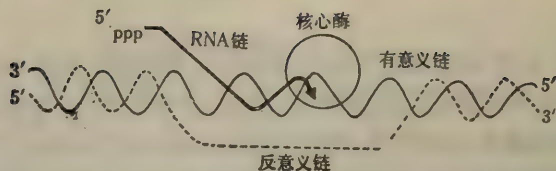
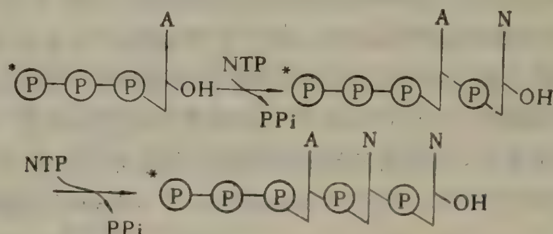


图 5-3 大肠杆菌 RNA 聚合酶的核心酶沿模板的有意义链移动合成 RNA

下列实验证明了转录的延伸方向是沿模板链(有意义链)的 3′ → 5′,生成 5′ → 3′RNA 产物。在 RNA 生物合成系统中加入 γ-^{32}P-ATP 或 γ-^{32}P-GTP, 观察 ^{32}P 参入到 RNA 中的动力学表明,合成早期 ^{32}P 参入最多,然后随着时间延长而下降;同时,已被 ^{32}P 标记的 RNA 总剂量并不因反应液中加入大量非同位素 ATP 或 GTP 而有所影响。因此,^{32}P 进入 RNA 是在合成的起始而不是合成的最后。由于 DNA 模板与 RNA 产物具反平行关系,所以转录时 RNA 聚合酶沿 DNA 的有意义链 3′ → 5′ 移动,而从 5′ → 3′ 方向合成 RNA。以 γ-^{32}P-ATP 为底物合成为例:



当用 3′-脱氧 ATP 参入到 RNA 链中时,将使链的延伸停止,这也说明延伸方向是 5′ → 3′。

一般来说,RNA 链生长的速度在体内和体外是相似的,有人估计,大肠杆菌在 37℃ 时,延伸速度是 25—50 个核苷酸/秒。

RNA 链延伸时所合成的多核苷酸链并不全保留在模板上,对一条 RNA 合成链而言,在任何时候只有一个连接点,该点沿 DNA 链移动,连续转录出相同的 RNA 产物,其长度随远离起始点而增加。图 5-4 是两栖类细胞 rRNA 基因转录过程的电镜图。mRNA 生物合成的电镜照片见图 5-21,从图中可以看到m RNA 上已经附着许多核糖体。

转录过程与复制有许多相似之处,但由于尚未发现转录有像复制那样的校正系统,所以转录较之复制出现错误的频率应是很低的,即转录应具有高度的忠实性。

转录的终止

在 DNA 分子上有终止转录的特殊信号,这些信号中有一些被 RNA 聚合酶自身识别,而另一些则需要一个蛋白因子 ρ 的帮助。ρ 是一个分子量为 20 万的四聚体,它能与 RNA 聚合酶结合,但不是酶的组分。体外实验表明,ρ 存在与否对 RNA 链合成的长短甚有影响,以 fd 噬菌体 DNA 为例,在无 ρ 因子时,合成的 RNA 产物为 26S,在合成开始后不同时间加入 ρ 因子,将分别得到 13S,17S 和 23SRNA,这说明模板上至少有三个为 ρ 识别的终止信号(图 5-5)。

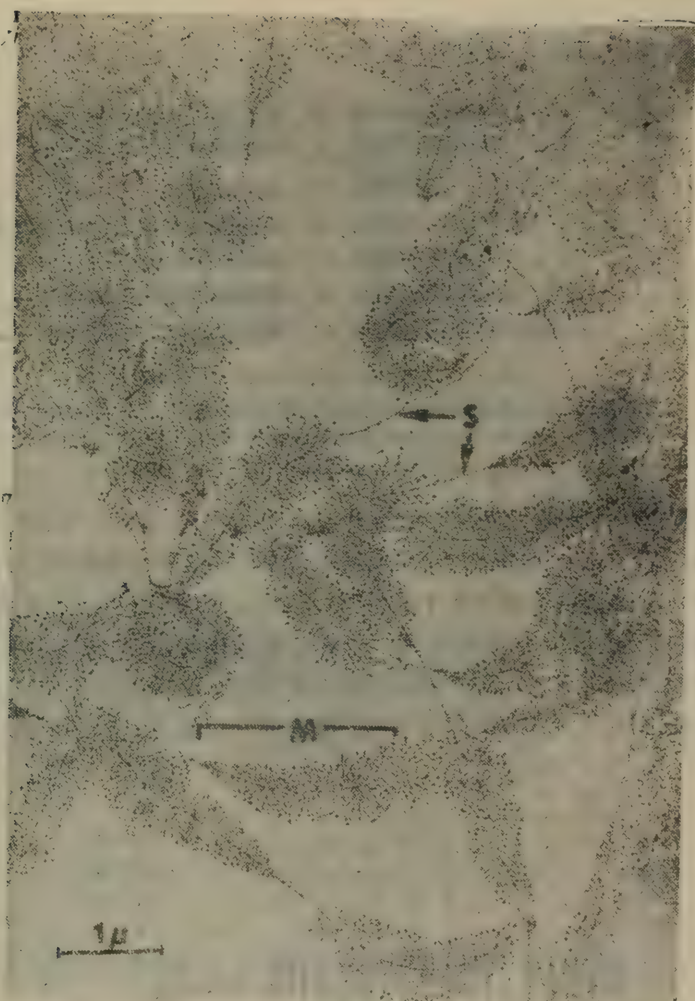


图 5-4　两栖类 rRNA 基因转录过程电镜图

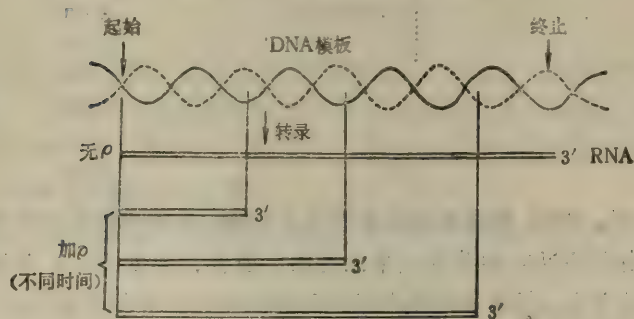


图 5-5　ρ因子对转录终止的影响
(图示 DNA 上有三个ρ作用点)

现在已有许多 RNA 转录产物 3′ 端的顺序 和 DNA 分子上终止区域的结构搞清楚了。虽然其一级结构没有什么共同之处，但发现终止区域在接近终点的部位都有一个二元对称结构，亦即回文结构，紧随着有数个 T 连在一起。 RNA 终止通常发生在二元对称结构中心以外 20±4 个核苷酸的位置。图 5-6 列出部分终止子反意义链的序列，具有明

显的茎环结构特征，同时后随数个 T。

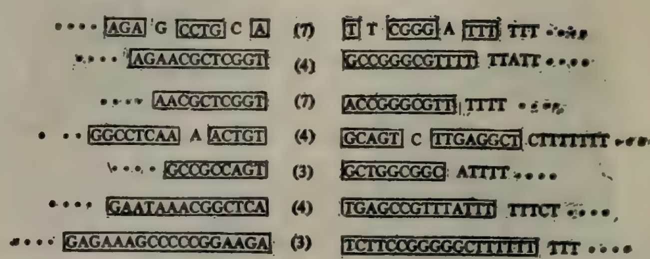
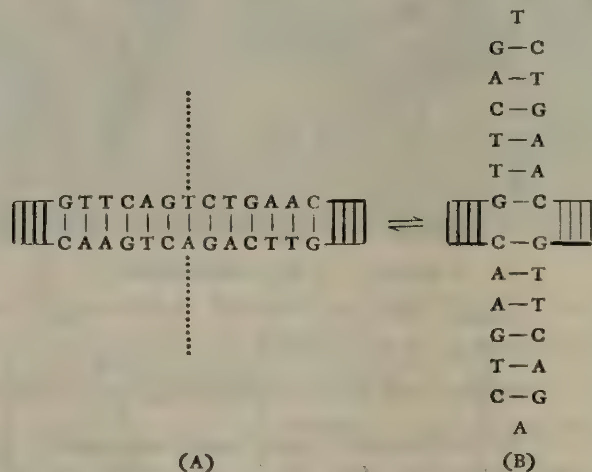


图 5-6　部分操纵子噬菌体的终止子结构特征

方框表示二元对称结构的碱基配对部分，组成茎环结构的茎，其中包括 G-T 对；括弧内为构成茎环结构的环的核苷酸数；在所有例子中下游都紧随数个 T。

摘自 *Ann. Rev. Genetics*, 1979, **13**, 319—353。

DNA 终止子部位的二元对称结构可以解释为什么 RNA 聚合酶在该处停止延伸：(1)早在 1966 年 Gierer 提出，二元对称结构(A)可以形成十字结构(B)，同时两者之间存在一个平衡；(B)结构由于少了几个碱基对，不如(A)结构稳定，但(B)形式结构可以为酶蛋白所识别并与之反应，从而防碍 RNA 聚合酶的移动，因而使转录延伸停止。(2)二元对称结构转录产生的 RNA 部分可以形成茎环结构，这样的结构可使 RNA 聚合酶发生构象变化，使之从“延伸”构象变为“终止”构象而终止转录，图 5-7 是 RNA 聚合酶终止



转录的一个模型。RNA 聚合酶延伸至终止部位时，转录产物仍然与有意义链模板氢键结合(图 5-7a)；由于 RNA 中茎的一半已游离出模板，通过 RNA 分子内相互作用而形成茎环结构(图 5-7b)，在这个区域中二元对称结构的变异将对茎环的形成发生重要影响；茎环的生成促使该处 DNA 链退火恢复螺旋结构，使茎环位于螺旋之外(图 5-7c)，这时 RNA 聚合酶就从“延伸”构象转变为“终止”构象。又由于茎环后一般为数个连续的 U，rU-dA 配对的 RNA-DNA 杂交体很不稳定而释放出转录产物。可以预料，减弱茎环或加强茎环后顺序与模板的相互作用都可能使转录“读过头”。

不同模板终止子结构或其生成的 RNA 茎环的大小以及稳定性是不一样的。例如，噬菌体 λ 6S RNA 的茎是很长的，且富含 G-C 对，而依赖于 ρ 的噬菌体 λ cro 基因转录

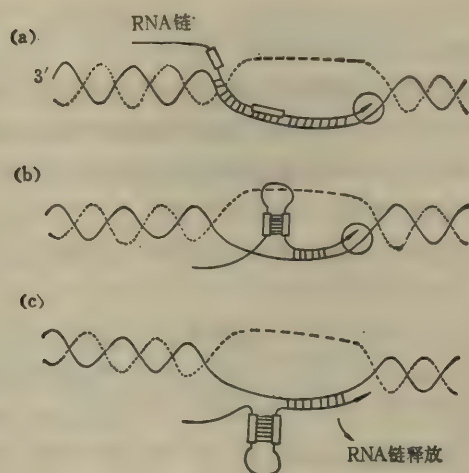
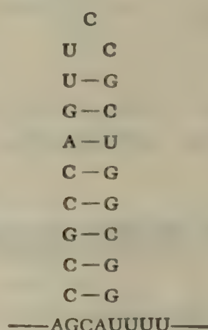


图 5-7　RNA 聚合酶终止转录模式

产物的末端茎最短，同时最不稳定。

最近对 trp 操纵子终止信号的研究表明，在操纵子最末一个结构基因之后存在两个终止信号，第一个信号结构为：



具有通用的结构特点，在数个U之前存在茎环结构，被认为这是一个体内终止信号。但如在体外加入 ρ 因子，则 RNA 聚合酶在离第一个终止信号下游 250 核苷酸处终止转录，该处顺序并无上述结构特征，但 A-U 丰富——UUAAUUUGACUAAUUCUCAUU——，因此可能存在另一未被认识的终止机制。

在依赖 ρ 的终止点处结构不够稳定，RNA 聚合酶只是在此暂停但不终止，需要 ρ 的帮助才能停止转录。ρ 具有二种活性：(1)终止转录；(2)依赖于 RNA 的核苷三磷酸酶（NTPase）活性。同时多种试验表明，这二种活性是同一蛋白质所具有，因此提出一个 ρ 终止转录的模型：ρ 因子开始附着在新生 RNA 上，随之沿 5′—3′ 方向朝 RNA 聚合酶移动，ATP 或其他核苷三磷酸水解提供这种移动的能量，最后，ρ 接触到在依赖 ρ 的终止点暂停的 RNA 聚合酶，与之反应，使新生的 RNA 链释放出来。从这个模型可以清楚地看到，影响 ρ 对 RNA 的结合或沿 RNA 移动的因素将抑制转录的终止。在原核生物中转录与翻译往往是偶联在一起的，翻译的同时进行意味着核糖体遮盖新生 RNA，而阻止 ρ 的结合，从而抑制转录的终止；翻译一旦停止，核糖体与 RNA 脱离，因而

出现裸露的 RNA，ρ 发生作用而使转录停止，这就是操纵子末端终止的机制。

DNA 模板上终止转录的信号叫终止子（terminator），它位于结构基因之后，如前所述。另外，在某些操纵子的结构基因之前也存在类似的结构区域，但只是在特殊环境下才表现为终止信号，通过结构的互变而行使调节基因表达的功能，这个部位叫做衰减子(attenuator)。Trp、Phe、Thr 和 His 等氨基酸操纵子都存在衰减子，其中以对 trp 操纵子的研究最为详细，详见本章第四节。

图 5-8 小结原核生物的转录过程，其中包括许多分子循环。

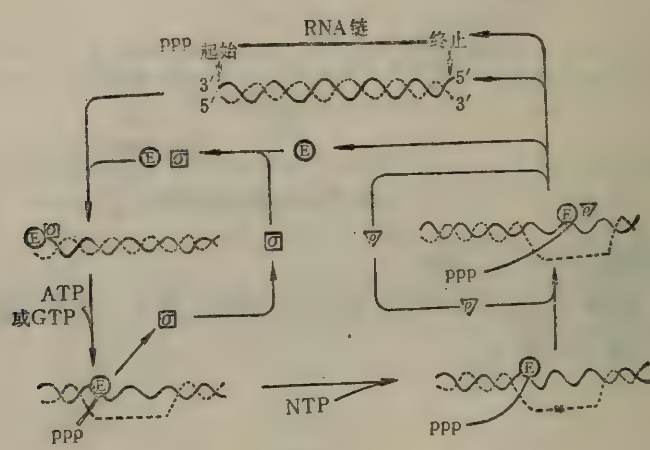


图 5-8 原核生物转录过程示意图

E 为 RNA 聚合酶的核心酶，σ 为聚合酶亚基，ρ 为与转录终止有关的蛋白因子。

三、转录的抑制剂

包括某些抗生素在内的转录抑制剂是研究转录机制的常用工具，同时，有的抑制剂也被用作治疗疾病特别是恶性病的药物。按照作用性质转录抑制剂一般可分为二大类：一类是通过与 DNA 结合而改变模板的功能；另一类则与 RNA 聚合酶结合而影响其活性。

表 5-5 与 DNA 模板结合的转录抑制剂

抑制剂	抑制机制	抑制效应
放线菌素D	插入二个 dG-dC 对之间	RNA 及 DNA 链延伸
纺綞菌素	通过氢键与静电相互作用与 DNA 结合	DNA 与 RNA 链生长
远霉素	同上	同上
蒽环糖苷类	插入 DNA 的碱基对之间	RNA 及 DNA 链延伸
吖啶类	插入 DNA 中	RNA 链起始，DNA 框码突变
溴乙锭	插入二个连续的 DNA 碱基对间	DNA 和 RNA 合成
黄曲霉素		DNA 及 RNA 合成
2-乙酰氨基芴		DNA 及 RNA 合成
蒽霉素		DNA 及 RNA 合成

与 DNA 模板作用的转录抑制剂

表 5-5 列出一些与 DNA 模板作用的转录抑制剂，除个别外，这类抑制剂都与 DNA 结合生成非共价复合物，从而削弱其模板活性。

放线菌素D是一种广泛使用和研究得最多的核酸合成抑制剂，它含有一个吩噁嗪酮(phenoxa zone)的平面发色团，通过二个侧链羧基各与一个环五肽联接。通过对放线菌素D与 dG 复合物(1：2)的晶体研究，提出了放线菌素D与 DNA 结合的分子模型：吩噁嗪酮环象一把刀子插入 DNA 双链中两对 dG-dC 核苷酸对之间，两个环五肽各位于吩噁嗪酮平面的两侧，与二条链上的 dG 的 2-氨基形成特殊氢键（图 5-9）。放线菌素D在低浓度（1mM）时，对大肠杆菌 RNA 聚合酶与 DNA 结合或 RNA 合成的起始影响很小，而选择性地阻止 RNA 链的延伸；在高浓度（10mM）时，它也抑制 RNA 合成的起始。放线菌素D主要抑制 RNA 的生物合成，对 DNA 合成也有小的影响。



放线菌素D

Sar 为 N-甲基甘氨酸即肌氨酸，MeVal 代表 N-甲基缬氨酸。Thr 的羟基与 MeVal 的羧基连接，形成二个相同的环五肽，它们分别位于吩噁嗪酮环平面的两边(图中未示出)。

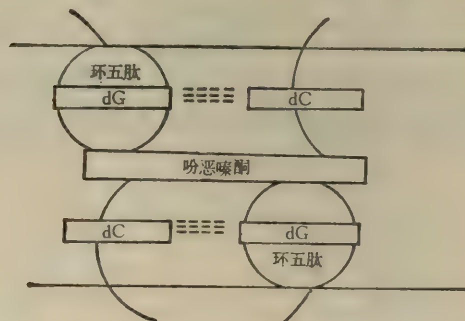
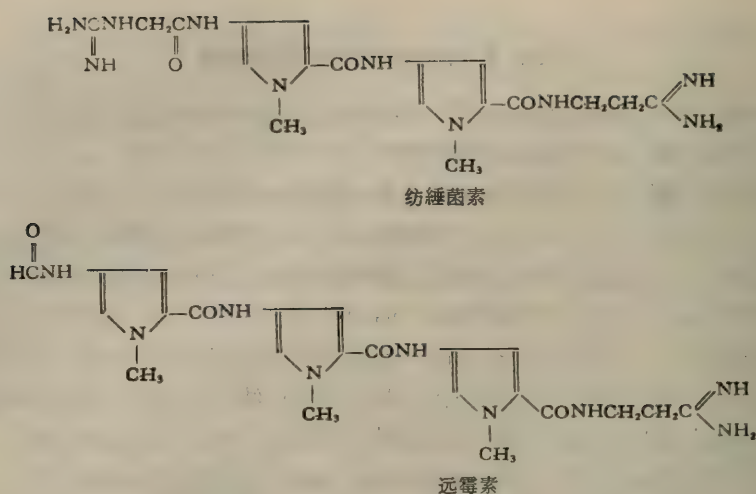


图 5-9　放线菌素D与 DNA 的结合模型

纺锤菌素（netropsin）和远霉素（distamycin）是一类寡肽抗生素，与双股 DNA 特别是多聚 dAT、多聚 dIC 的碱基对形成非常稳定的复合物。与放线菌素D的插入结合不同，而是依赖于氢键和静电相互作用。

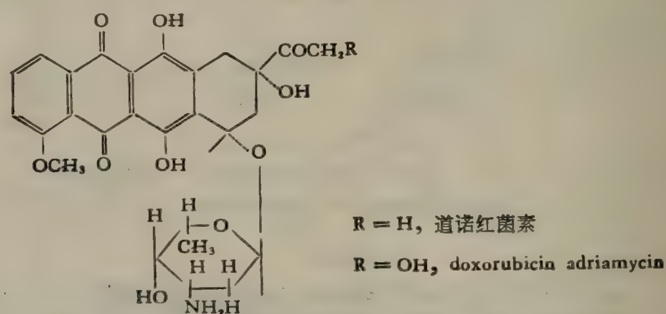


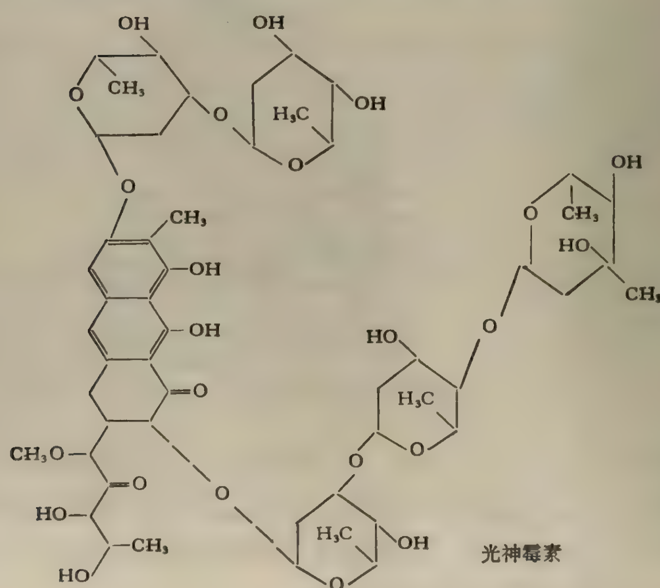
纺锤菌素

HCNH
O
CONH
N
CH₃
CONH
N
CH₃
CONHCH₂CH₂C
NH
NH₂
N
CH₃

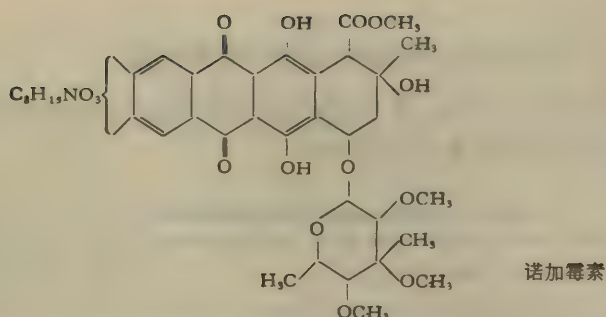
远霉素

蒽环糖苷类（anthracycline glycosides）包括道诺红菌素（daunorubicin）家族、光神霉素（mithramycin）和诺加霉素（nogalamycin）等，其中的四环结构可插入碱基对之间，从而抑制 DNA 及 RNA 链的生长。



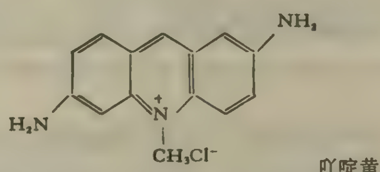


光神霉素



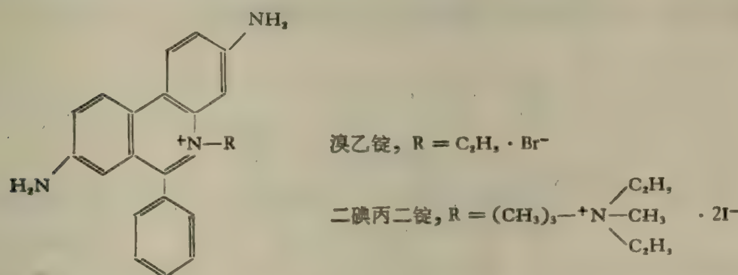
诺加霉素

吖啶类（acridines）染料与 DNA 结合能引起框码突变(见第七章)，此外，它也能抑制 RNA 链的起始以及质粒复制。这类化合物有原黄素（proflavine，结构见第七章)、吖啶黄（acriflavine）等。

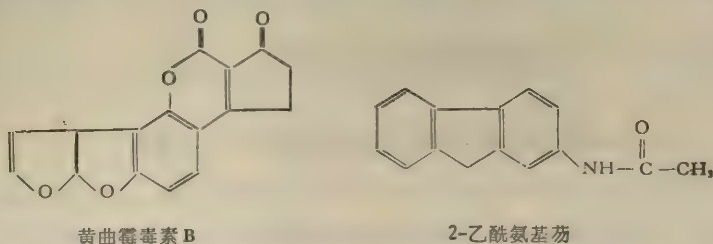


吖啶黄

溴乙锭也能插入 DNA 碱基对之间，是检测 DNA（以及某些 RNA）的一个高敏感度的萤光试剂。与 DNA 结合后，抑制 DNA 与 RNA 两者的生物合成。相类似的抑制剂还有二碘丙二锭（propidium diiodide)。



黄曲霉毒素（aflatoxin）和 2-乙酰氨基芴（2-acetyl-aminofluorecene）作用于 DNA，同时抑制复制与转录。黄曲霉毒素从生长在花生上的黄曲霉菌（*Aspergillus flavus*）中分离得到，2-乙酰氨基芴则是人工合成化合物，两者均为烈性致癌物，能诱发肝癌。



黄曲霉毒素 B　　　　2-乙酰氨基芴

蒽霉素（anthramycin）也是一种与 DNA 共价结合的抑制剂，抑制 DNA 和 RNA 的生物合成。

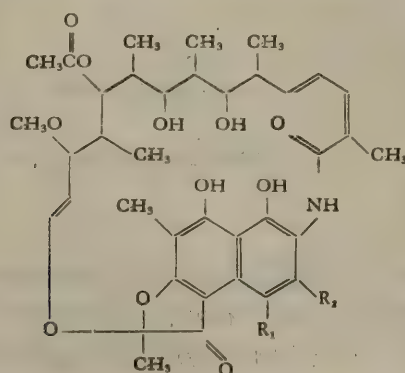
与 RNA 聚合酶作用的转录抑制剂

表 5-6 列出几个主要的与 RNA 聚合酶作用的转录抑制剂，这些抑制剂对于研究转录机制和 RNA 聚合酶的性质起着很大的作用。

表 5-6 与 RNA 聚合酶作用的转录抑制剂

抑制剂	作用方式	抑制效应
利福平和曲张链菌素	与微生物和线粒体 RNA 聚合酶的 β 亚基非共价结合，不防碍酶与 DNA 结合，而阻碍第一个 5'-核苷三磷酸进入	阻断 RNA 链的起始
利链菌素	与细菌 RNA 聚合酶的 β 亚基结合	优先阻断 RNA 链的延伸
α-鹅膏蕈碱	与哺乳动物细胞核 RNA 聚合酶 II 优先结合；不影响细菌、线粒体和叶绿体的 RNA 聚合酶	专一抑制哺乳动物 mRNA 的合成。高浓度的 α-鹅膏蕈碱也抑制其它 RNA 的合成

利福平 (rifampicin) 是利福霉素的衍生物，利福霉素是链霉菌属产生的一种抗生素，它有效地对付抗酸 (acid-fast) 菌和革兰氏阳性菌。利福平治疗结核病特别有效。利福平是细菌 RNA 聚合酶的非常有效的抑制剂，$2\times10^{-8}M$ 浓度可抑制 50% 活性。抑制机制是与酶的 β 亚基结合，而防碍了 5'-嘌呤核苷三磷酸的进入，结果阻断 RNA 链的起始。利福平并不影响酶与 DNA 的结合。利福平对真核 RNA 聚合酶无作用，然而在某些特定情况下，真核 RNA 聚合酶可被一些利福平衍生物所抑制。

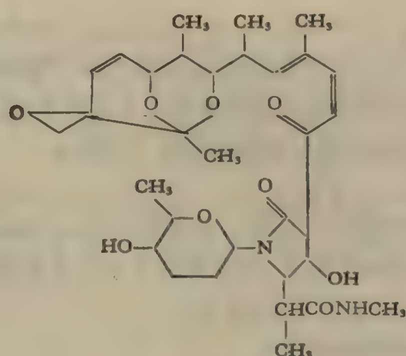


利福霉素 B：$R_1 = -OCH_2COOH$，$R_2 = -H$。

利福平：$R_1 = -OH$，$R_2 = -CH_2=N-N\langle\text{哌嗪环}\rangle N-CH_3$。

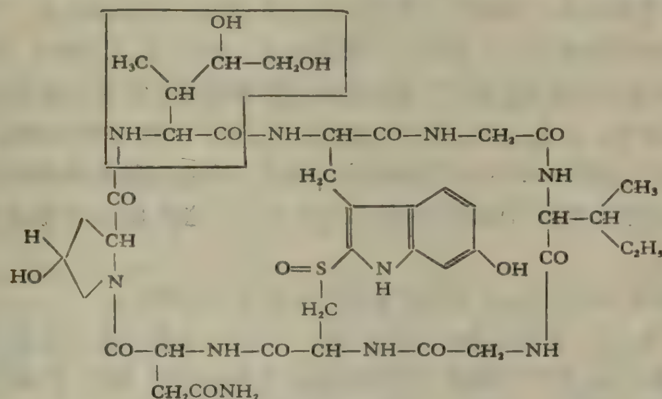
曲张链菌素 (streptovaricin) 具有与利福平同一专一性和作用方式，来源也相同。

利链菌素 (streptolydigin) 与利福平不同，虽然作用部位仍然是细菌 RNA 聚合酶的 β 亚基，但却优先阻断 RNA 链的延伸。



利链菌素

α-鹅膏蕈碱（*α*-amanitin）是从毒麻菇 *Amanita phalloides* 分离出来的一个剧毒物质，为一环八肽化合物，由 Gly 二分子和 L-Ile、L-二羟 Ile、L-Asn、L-羟 Pro、L-羟 Trp 和氧化 Cys 各一分子组成，其中 Trp 残基中心与氧化的 Cys 巯基相联接。



α-鹅膏蕈碱

如果 *α*-鹅膏蕈碱中的二羟 Ile 残基(见图中方框部分)被非羟化的 Leu 代替，则产生完全无毒性的 Amanullin。从毒麻菇中也分离出几个毒性七肽。

α-鹅膏蕈碱完全阻断真核 RNA 聚合酶 II 以及 III 产生的转录作用，特别抑制 mRNA 的生物合成，从而也抑制了整个细胞的蛋白质生物合成。但对细菌、线粒体和叶绿体的 RNA 聚合酶没有影响。

此外，RNA 聚合酶的非特异抑制剂有：多核苷酸、肝素、聚乙二醇磷酸、核苷三磷酸类似物，如 3′-脱氧 ATP，3′-氨基-3′-脱氧 ATP 和 2′-O-甲基 ATP 等。

四、mRNA 的转录与加工

mRNA 概念的提出和 mRNA 的发现

mRNA 概念是 Jacob 和 Monod 于 1961 年提出的。在这之前，已有许多研究工作者在思考基因（DNA）、RNA 与蛋白质生物合成之间的关系。在真核细胞中，遗传信息贮存于细胞核的 DNA 之中，而蛋白质合成部位却在细胞质，那么在这两者之间必定有一个或一类化学中间物存在，它能把 DNA 上的信息传递蛋白质。该化学中间物曾被称

作"结构信使"或简称"信使"。Jacob 和 Monod 在研究大肠杆菌诱导酶的形成时发现，添加诱导物在数分钟内诱导酶可以上千倍地增加，而移去诱导物则酶合成迅速停止。因此，他们设想这些大量的酶在稳定的模板上直接形成是不可想象的，而必然存在一个短命的中间物即信使作为合成酶的模板，这个信使应具有下列性质：

(1) 信使应是多核苷酸；

(2) 信使的碱基组成应该反映于 DNA 所特有的碱基组成；

(3) 由于生物体内蛋白质种类繁多，故信使的大小应是很不均一的，根据三个核苷酸编码一个氨基酸计算，信使的分子量至少 50 万；

(4) 信使应能短暂地与核糖体结合；

(5) 信使的合成与分解应该是迅速的。

当时已经知道的二类 RNA (tRNA 于 1957 年发现，rRNA 在 1959 年分离)都不具有这些性质。第一，rRNA 和 tRNA 的碱基组成与 DNA 相差甚大，它们如何接受来自 DNA 的信息？第二，rRNA 和 tRNA 分子相对比较均一(见第三章)，何以解释生物体内存在千差万异的蛋白质种类？第三，这两类 RNA 代谢率太低，不能说明生物体具有高速度合成蛋白质的能力；第四，不同生物体 rRNA 以及 tRNA 碱基组成彼此相似，具有很大的同源和亲缘性(见第三章)，如果它们是蛋白质合成的模板的话，很难说明蛋白质的种族差异性。因此 Jacob 和 Monod 认为，除了 rRNA 和 tRNA 以外，必定还有一类 RNA 存在，这类 RNA 必需满足上述要求，在蛋白质合成中决定氨基酸的排列顺序，把从 DNA 分子得到的信息传递给蛋白质分子。这类 RNA 就被定名为信使 RNA (mRNA)。

mRNA 假说提出后不久，许多实验室立即证实了它的存在。

用噬菌体 T2、T7 感染大肠杆菌后，虽然 RNA 总量没有什么变化，但约有 1% RNA 是新合成的，而且更新率很快，同时发现这类新合成的 RNA 的碱基组成和大肠杆菌原来有的 RNA 不同，而与噬菌体 DNA 则甚类似(表 5-7)。

表 5-7 大肠杆菌经 T 噬菌体感染后新生成的 RNA 的碱基组成%

核酸种类	A	G	C	T	U
大肠杆菌 RNA	23	32	23		22
T2DNA	32	18	17*	32	
T2 感染后生成的 RNA	32	18	18		32
T7DNA	26	23.8	23.6	26.6	
T7 感染后生成的 RNA	26.7	21.8	23.7		27.8

* 为 5-羟甲基胞嘧啶。

在大肠杆菌或用噬菌体感染后的大肠杆菌的培养中，加入 ^{14}C 或 ^{32}P 同位素标记的化合物进行短时间培养，经密度梯度超离心分离。都能得到含高放射性的新峰，它既不是 4S 的 tRNA，也不是 16S 和 23S 的 rRNA，而是一类新 RNA。同时进一步发现，这个新的放射性高峰可由于在非放射物中较长时间培养(如 37℃ 保温 15 分钟)而消失，说明这类 RNA 含量既少，代谢很快。虽然这类放射性物质主要和核糖体结合在一起，但其

沉降常数不是 rRNA，并且可用电泳将其分离出来。

1961 年 Spiegelman 发展了一种新的杂交技术，为 mRNA 的存在提供更有说服力的证据。用 ^{32}P 标记经 T2 噬菌体感染大肠杆菌后生成的 RNA 即 T2 mRNA，同时制备 ^{3}H 标记的 T2 DNA，将 T2 mRNA 与 T2 DNA 混合物加热到 100℃。然后缓慢冷却到室温，冷却混合物进行密度梯度超离心，分析结果见图 5-10。 ^{32}P 和 ^{3}H 双重标记的杂交体的存在表明，T2 mRNA 不但其组成与 T2DNA 相似(见表 5-7)，而且两者的核苷酸排列顺序也有对应的互补关系。

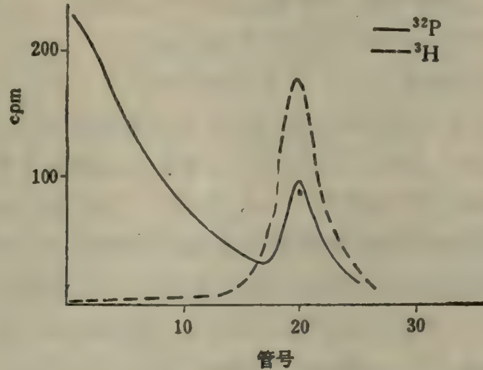


图 5-10　T2 噬菌体感染大肠杆菌生成的 RNA 与 T2 DNA 的杂交实验

不久，采用双标记法(^{3}H 标记 DNA 链，^{32}P 标记 RNA 链)也可在体内找到这种杂交产物。用硝基纤维过滤法能够更简便而快速地分离鉴定杂交分子。这样，经过许多实验室的努力，完全证实了 Jacob 和 Monod 预言的 mRNA 的存在。

稍后，在真核生物中也发现 mRNA。真核生物 mRNA 中 3′ 端都存在有多聚 A 片段，为分离鉴定带来了很大的方便(见第十一章)。鉴定 mRNA 的简便方法是体外无细胞翻译系统，常用的有麦胚无细胞系统，兔网织红细胞裂解液系统等。

原核生物 mRNA 的转录

原核生物 mRNA 转录有以下几个特点： (1)多顺反子转录，以操纵子作为转录单位；(2)转录与翻译偶联在一起，往往 mRNA 还未转录完全，蛋白质合成就已开始；而且，翻译成为调节转录的一个因素；(3)除个别 mRNA 外，一般原核生物 mRNA 不需要加工；(4)原核生物 mRNA 的寿命比较短，大肠杆菌 mRNA 的半寿期一般只有几分钟或者更短。

I. 原核生物 DNA 分子上的转录信号——操纵子及其调控区。

在原核基因组织中，几个相关的蛋白质或几个相关的 RNA 的基因常被组织为一个转录单位。包括几个蛋白质基因的转录单位被转录为一个多顺反子 mRNA，产物除编码几个蛋白质的序列外，也包括位于分子两端或顺反子间的非编码区，非编码区是 mRNA 的结构特征之一(见第三章)，它不翻译，但转录后这部分并不被加工释放。

1961 年 Jacob 和 Monod 在提出 mRNA 概念的同时，也提出操纵子学说，为原核生物转录及其调控方式设计了一个简单的模型。虽然自此以后，大量研究结果表明，他们的模型是相对简单化了，但所有原核生物的基因表达都可以用操纵子作为转录单位予以描

述。操纵子包括数个相关的结构基因及其调控区，是一个基因表达的协同单位。调控区主要可分为三个部位——操纵基因、启动子和降解物基因活化蛋白（CAP）连接部位，它们之间通常相互重叠。启动子是 RNA 聚合酶的结合部位，结合后启动 RNA 合成；操纵基因是操纵子以外的调节基因生成的阻遏蛋白的结合部位，二者结合后，转录即被阻断或大大地减慢速度，这叫转录的“负调节”；CAP 由分子量为 45,000 的两个亚基组成，也叫 cAMP 受体蛋白，cAMP 促进 CAP 与 DNA 结合而加快转录速度，可能是由于它们在结合点使 DNA 局部变性，增强 RNA 聚合酶对启动分子的结合而促进转录。这叫转录的“正调节”。此外，在某些操纵子的调控区中存在衰减子，这个部位不同于结构基因下游的终止子，它既是一个终止信号，也是一个调节信号。

对二类操纵子进行了详细的研究：一是可诱导操纵子，另一为可阻遏操纵子。可诱导操纵子指在正常的情况下不表达或仅在低水平上表达的转录单位，如果环境发生某些变化，则操纵子开放，转录 mRNA，从而合成诱导酶以适应细胞的需要。许多负责糖分解代谢的操纵子属于这种类型，如乳糖（lac）操纵子、半乳糖（gal）操纵子和阿拉伯糖（ara）操纵子等。可阻遏操纵子则主要是调控氨基酸过剩的，氨基酸过多，可以通过两条途径调节，其一是负反馈以抑制合成该氨基酸所需要的一系列酶中的关键酶，另一种方式是关闭基因使酶不能合成，显然后者是最有效的。可阻遏操纵子就是负责氨基酸合成代谢调控的操纵子，如色氨酸（trp）操纵子、组氨酸（his）操纵子等。

为了简便，在下列分述的操纵子中，统一采用以下符号和代号：i，调节基因；p，启动子；o，操纵基因；a，衰减子；结构基因则以相应大写符号表示。●，RNA 聚合酶；○，阻遏蛋白；△，CAP-cAMP（CAP 与 cAMP 的复合物）；～～～，mRNA； ꝏ蛋白质(酶)。

lac 操纵子是研究得最早而又最详细的，其基因结构排列在第二章已介绍过了，图 5-11 示出 lac 操纵子的阻遏（正常情况下，无诱导物存在）和诱导（加入乳糖或其类似物）过程。在 lac 操纵子的调控区中，各信号蛋白结合部位的核苷酸顺序已经测出（见图 5-12），其中启动子的序列具有通常的保守序列（见第二节，转录的起始），CAP-cAMP 结合部位和操纵基因则具有紧密的回文结构，见图 5-12 中的方框部分。关于 CAP-cAMP 的结合位点及阻遏蛋白与操纵基因的相互作用问题，分别参见第九章和第十章。

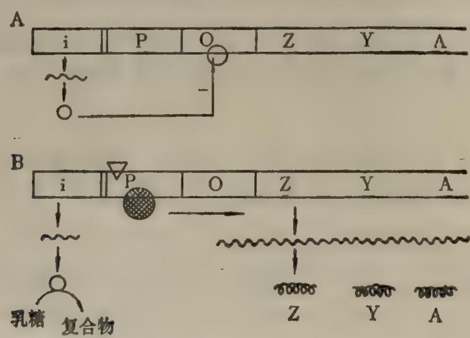


图 5-11 lac 操纵子的阻遏（A）与诱导（B）

gal 操纵子　在第二章已提到，它含有三个结构基因 E、T、K 和一个 92 核苷酸

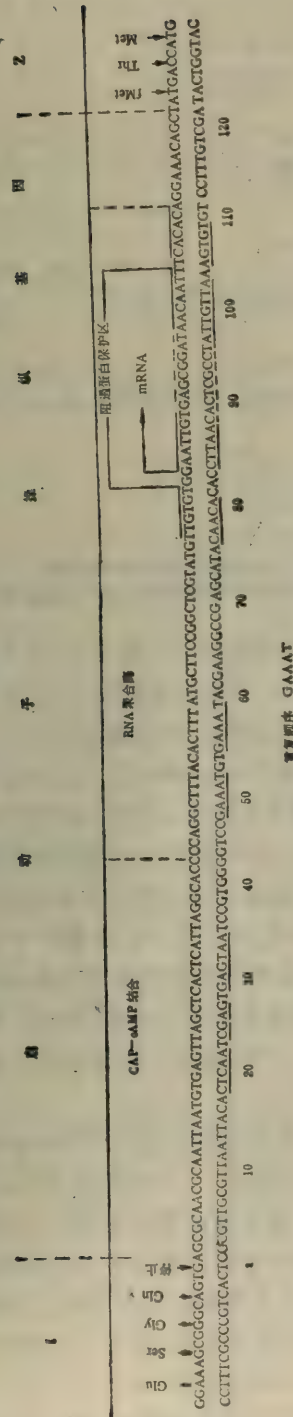


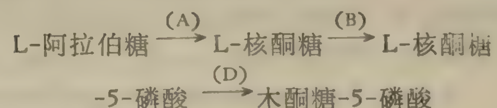
图 5-12 lac 操纵子调控区的结构特征

长的调控区。 E、T、K 分别编码三个酶，这些酶使半乳糖转变为 G-1-P 和 UDPG。虽然 gal 操纵子与 lac 操纵子的表达调控方式基本相似，但体内实验观察到有明显的不同：(1)虽然 CAP-cAMP 有明显的激活效应，但 gal 操纵子不象 lac 操纵子那样依赖于降解物的活化；(2)当操纵子被阻遏，仍出现低基础的酶合成，合成水平远高于其他操纵子。 图 5-13 列出 gal 操纵子中调控区的排列情况。当与半乳糖代谢有关的酶被诱导和存在 CAP-cAMP 时，操纵子在 1 处开始转录。整个调控区位于左侧，核苷酸数目用负数表示，启动子部位在 1 至 −60，再左侧 −60 至 −90 核苷酸为 CAP-cAMP 结合部位，操纵基因则与 CAP-cAMP 部位重叠，−60 至 −66 核苷酸为阻遏蛋白结合处。

图 5-13 gal 操纵子

由于这样的结构特点，阻遏蛋白的存在首先干扰降解物活化过程，但不完全破坏 RNA 聚合酶与启动子相互作用，故也能进行低水平转录。 另外，在缺乏 CAP-cAMP 的情况下，转录亦能缓慢地进行，但转录的起始点与存在 CAP-cAMP 时不同，后者从 1 位开始，前者则在 −5 位启动(见图 5-13)，同时不被阻遏蛋白所影响。

ara 操纵子 含有三个结构基因 B、 A 和 D，能表达三个用于阿拉伯糖代谢的相关的酶——核酮糖激酶 (B)、阿拉伯糖异构酶 (A) 和核酮糖-5-磷酸差向异构酶(D)，催化下列反应：



$$\text{L-阿拉伯糖} \xrightarrow{(A)} \text{L-核酮糖} \xrightarrow{(B)} \text{L-核酮糖-5-磷酸} \xrightarrow{(D)} \text{木酮糖-5-磷酸}$$

调节基因产生的阻遏蛋白存在二种别构形式 P_1 和 P_2，没有阿拉伯糖诱导物，主要以 P_1 形式存

在，作为阻遏蛋白与操纵基因结合，实现负调节(见图 5-14A)；有阿拉伯糖时，P_1 转变为 P_2，能作为激活剂与启动子结合，实现正调节(见图 5-14B)。此外 CAP-cAMP 和 ppGpp (见第九章)也都能促进 ara 操纵子基因表达，但结合部位不很清楚。

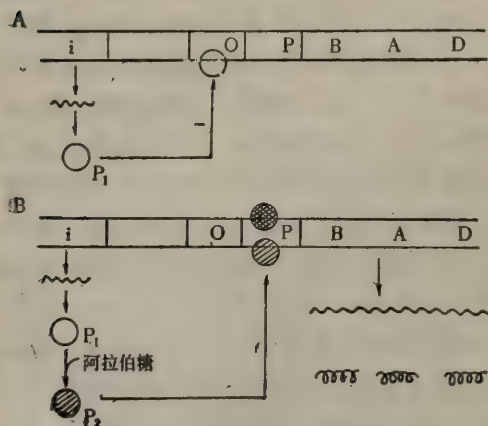


图 5-14 ara 操纵子的调控

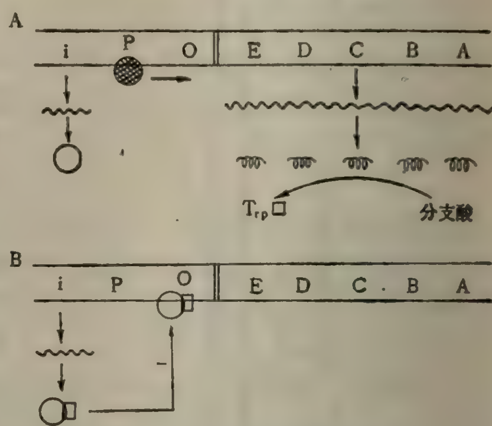


图 5-15 trp 操纵子的早期调控模型
trp 作为共阻遏物与阻遏蛋白结合后抑制转录

trp 操纵子　　它是目前研究得较为详细的一个可阻遏操纵子。图 5-15 是早期建立的 trp 操纵子结构和调控模型，它包括五个结构基因以及 5′ 旁侧的调控区。五个基因产物组成的酶系可以催化从分支酸合成 trp(图 5-15A)，trp 过剩时，则阻遏转录进行。按照 Jacob 和 Monod 的操纵子概念，提出 trp 是一个共阻遏物 (corepressor)，当它与阻遏蛋白结合后，即与操纵基因连结而阻断转录(图 5-15B)。但是后来发现，有些实验结果并不能用上述模型解释，如于 trp-$tRNA^{Trp}$ 变异株中加入 Trp 并不减低 trp 操纵子的基因表达；加 Trp 于 Trp 饥饿细胞并不关闭转录的起始，而只抑制起始片段上转录的延伸，同时也有实验说明，在操纵子的起始片段中可能存在终止部位，因为在 trp 操纵子 5′ 端序列产生的 mRNA 中，相当于操纵子的前导序列 140 bp 的寡核苷酸多于来自其他部分的片段几倍。基于这些观察，人们提出另一种调控机制，叫做衰减作用 (attenuation)，即通过转录的终止(而不是转录的起始)对基因表达进行调控——Trp 缺少，转录进行；Trp 过剩，转录终止。

1981 年已全部测出 trp 操纵子中 7,000 多个 bp 的排列顺序，其中 6,800bp 构成五个结构基因，编码酶或其亚基，催化从分支酸经过五步反应合成 L-Trp(图 5-16)。trp 操纵子结构有以下几个特点：(1)在结构基因前有一段 162 核苷酸的前导序列(L)，能编码出一个 14 肽，其转录产物具有一奇特的二级结构，其中含有被称作衰减子(a)的转录终止的调节信号；(2)启动子与操纵基因大部分重叠；(3)除了 5′旁侧的启动子外，在操纵子内部存在另一弱的启动子(P_2)；(4)结构基因之后出现二个终止信号 (t 和 t′)，其中一个(t′)为 ρ 因子所识别(见第 147 页)。trp 操纵子的调控区排列见图 5-17。

trp 操纵子前导序列的转录产物具有下列奇特的可变二级结构(图 5-18)，图中方框内为典型的终止信号(见第二节)，结构左面包括前导肽羧端的 6 个氨基酸密码子及其终止密码子(见 I 旁所注)。

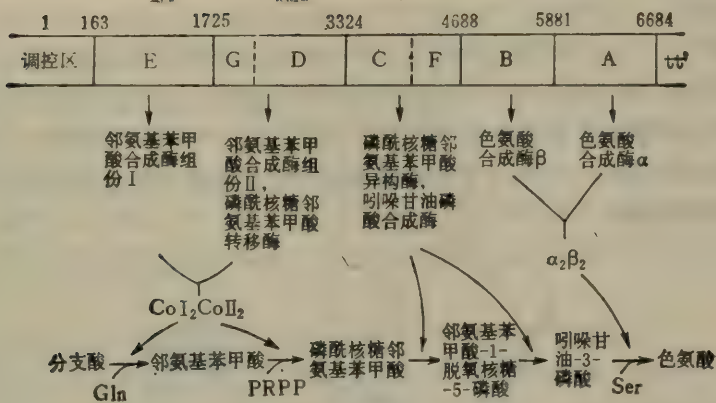


图 5-16　trp 操纵子结构基因排列及其表达产物

tt' 表示两个终止子，其顺序见第 147 页；数目字代表从启动子右侧起的核苷酸对数。

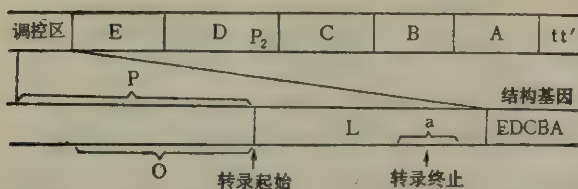


图 5-17　trp 操纵子的调控区排列

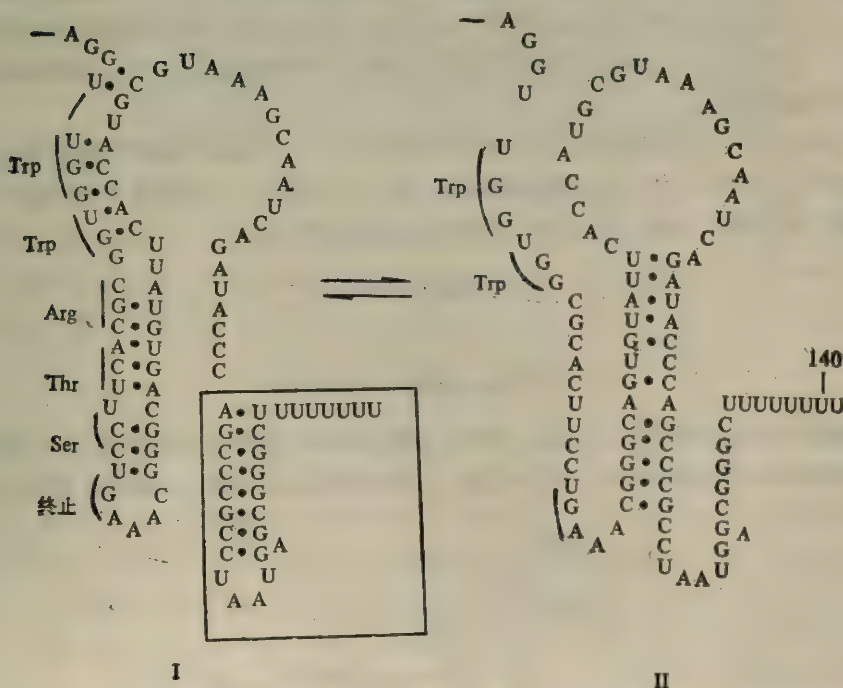


图 5-18　trp 操纵子前导序列的转录产物的二级结构图

在结构基因前存在这样结构的衰减子可以解释基因表达是如何通过转录终止进行调控的：由于原核生物转录与翻译是偶联在一起的，这种情况下，若(1)，前导肽不被翻译

时，I 型结构稳定，则转录在终止信号(方框处)停止；(2) Trp 缺少，则 Trp-tRNA^{Trp} 不能生成，翻译至 Trp 密码子处停止，这时，核糖体使 I 型结构破坏，从而不存在终止信号结构，得到类似 II 型的结构。RNA 聚合酶于是顺利通过进行转录以合成 Trp；(3) Trp 过量，当然还要存在正常量的 tRNA^{Trp} 以及 Trp-tRNA^{Trp} 合成酶，则前导 14 肽被翻译，这时，核糖体破坏 I 结构中左侧的氢键配对区而保留其终止信号，故转录也停止。三种情况的示意图见图 5-19。

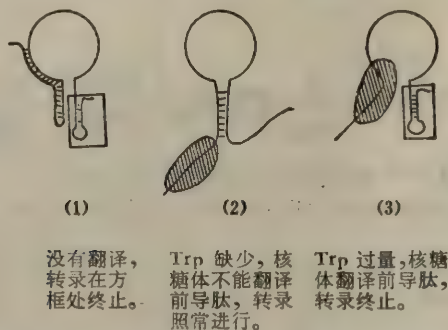
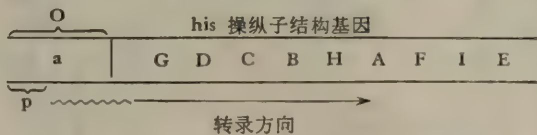


图 5-19　大肠杆菌 trp 操纵子中衰减作用模型

◍表示核糖体，核糖体大小见第六章，约能遮盖 16-20 核苷酸。

综上所述，在 trp 操纵子的基因表达中，受到 Trp-tRNA^{Trp} 的影响，而 tRNA^{Trp} 与 Trp-tRNA^{Trp} 的相对浓度则决定于细胞内 Trp 浓度、Trp-tRNA^{Trp} 合成酶的量和活性、tRNA^{Trp} 量、甚至最重要的是细胞内整个蛋白质合成的速度。通过阻遏作用进行转录调控(见图 5-15)不是 trp 操纵子的唯一机制，而衰减作用模型(图 5-19)，则能解释许多已经存在的实验结果。

his 操纵子　它具有类似 trp 操纵子的调控方式，实际上衰减作用概念最早是从沙门氏杆菌 his 操纵子研究中首先提出的。在 his 缺乏时，his 操纵子能产生 9 个酶催化从 ATP 和 PRPP 合成 His。his 操纵子中基因排列如下：



和其他操纵子有类似情况，his 操纵子的结构基因顺序与酶的作用顺序并不一致。his 操纵子没有一个专门的调节基因，而是由第一个结构基因 G 产物——磷酰核糖基-ATP 合

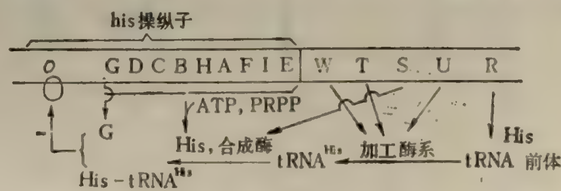


图 5-20　his 操纵子的负调控

R 是 tRNA^{His} 结构基因；U、W 为 tRNA 加工酶基因；T 为修饰酶基因，该酶催化 U→φ；S 表达产生 His-tRNA^{His} 合成酶。

成酶和 His-tRNA^{His} 一起组成有活性的阻遏物进行负调节，因此所有有关形成 His-tRNA^{His} 的酶基因的突变都直接影响 his 操纵子的表达(图 5-20)。

his 操纵子的调控区序列已经测定，其前导序列能编码一个 16 肽，其中含有连续的七个组氨酸残基。前导序列转录产生 mRNA 具有类似于 trp 前导 mRNA 的可变二级结构。His 缺少，转录通过结构基因；His 过量或缺乏对应于 His 密码子前任一氨基酸密码子的氨酰 tRNA，则前导 mRNA 上出现终止信号而使转录在衰减子处停止。但是，his 操纵子前导序列形成的二级结构较 trp 操纵子复杂，正常情况下出现四个长短不一的茎环区。

II. 原核生物转录与翻译偶联在一起，一般原核 mRNA 不需加工

从拍摄的原核生物转录过程的电镜图(如图 5-21)，可以观察到在新生的 mRNA 链上附着许多核糖体，这意味着转录尚未结束而翻译早已开始，新生 RNA 即为成熟 RNA，而不必经过进一步加工。而其他所有种类的细胞质 RNA 都是要进行加工的（见下文）。

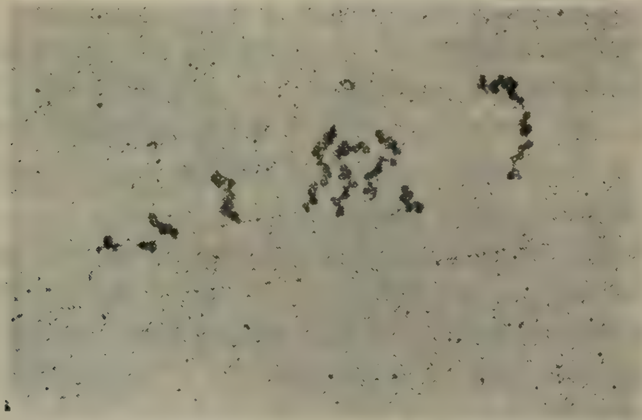


图 5-21　大肠杆菌转录与翻译的偶联

但是也有例外，如 T7 DNA 由一个启动子、5 个结构基因和一个终止子组成，被大肠杆菌 RNA 聚合酶全部转录得到较大的 RNA 前体，然后由加工酶 RNaseIII 剪切成五个成熟的 mRNA 分子。

真核生物 mRNA 的转录与加工

第三章已经讨论 mRNA 的结构问题。真核细胞的核、质分明，转录与翻译有一个生理上的屏障，与原核细胞有明显不同的特点。关于 mRNA 的转录问题，应该深入研究细胞核才能解决。由于真核细胞及其 DNA 的复杂性，花了很长时间才找到了追踪的线索，首先发现核内存在一类含量不算太少的 RNA，分子很不均一，无论在体内或体外标记实验，可以很快分离得到一系列标记化合物。同时证明了这些化合物不是 rRNA (18S 和 28S) 及其前体 (45S)。这类 RNA 就被称为不均一的细胞核 RNA(hnRNA，heterogeneous nuclear RNA)。hnRNA 一般要比细胞质 mRNA 大，大约从 10S 到 100S。哺乳动物 hnRNA 链长在 5,000 到 50,000 核苷酸之间。组织培养的哺乳动物细胞如 L 细胞，每一细胞约含 2pghnRNA，大约是细胞质 mRNA 的四倍。

hnRNA 半寿期变化很大，不同种类细胞有不同的半寿期。一般来说 hnRNA 半寿期

很短,比细胞质 mRNA 更不稳定。如L细胞 hnRNA 半寿期只有 0.4 小时,而其细胞质 mRNA 半寿期为 1—10 小时。海胆胚胎 hnRNA 半寿期仅为 7 分钟,而其细胞质 mRNA 半寿期为 3 小时。很早人们就设想 hnRNA 可能是 mRNA 的前体,或至少部分是 mRNA 前体。至今已有大量实验证据说明这一点,而序列分析的进展提供了肯定的证据。既然有前体,就必需经过加工才能产生有功能的成熟 mRNA。

真核生物 RNA 包括 mRNA 的转录具有与原核生物明显不同的特点。如转录起始信号有所不同(见第二节);真核转录单位往往是单顺反子,不象原核生物那样组织为操纵子;大多数蛋白基因以及个别 tRNA 基因(如酵母)、rRNA 基因(如果蝇)存在插入顺序,它把同一结构基因分成数段,同时在许多情况下,插入顺序链长比结构基因还长;随之而来的是,加工过程更为复杂等等。

hnRNA 是 mRNA 的前体

hnRNA 与 mRNA 存在非常密切的关系,表现为:

(1) 与 mRNA 相似,有类似于 DNA 的碱基组成,而与 rRNA 碱基组成很不相同。如 HeLa 细胞 hnRNA 含有 43—45% 的 G + C,非常类似于细胞 DNA (42—47% 的 G + C),而明显不同于 rRNA (> 60%的 G + C)。当然亦有少数情况 hnRNA 的碱基组成与 DNA 并不一样,在同一细胞核中不同染色体上,甚至同一染色体不同部分上合成的 hnRNA 碱基组成亦是不一样的。

(2) hnRNA 有与 mRNA 相同的 5′ 端。hnRNA 5′ 端具有 $m^7G^{5'}ppp^{5'}$ NmpNp-结构,与细胞质 mRNA 5′ 端相同。鼠L细胞 hnRNA 与 mRNA 具有十分相似的 5′ 端组成,如 5′ 端为 $m^7G^{5'}ppp^{5'}m^6$AmpG-的都占 20%,而 5′ 端为 $m^7G^{5'}ppp^{5'}$UmpG-的都不到 1%。

(3) hnRNA 的合成和细胞质 mRNA 合成一样,不为低剂量放线菌素D所抑制(低剂量放线菌素D能选择性地抑制 rRNA 的合成),但能为高剂量放线菌素D所抑制,这表明 hnRNA 与 mRNA 为同一类型 RNA 聚合酶所合成,这一结果亦支持关于在 hnRNA 与 mRNA 之间存在着前体与产物关系的设想。

(4) hnRNA 与细胞质 mRNA 存在着类似的加上多聚A过程。 大多数真核细胞 mRNA 中存在着 3′ 端多聚 A,而 hnRNA 分子中却只有 10—40% 含有多聚 A。哺乳动物 hnRNA 中多聚A长度平均为 150—200 个核苷酸,而其 mRNA 中多聚A长度为 20—200 个核苷酸。hnRNA 和 mRNA3′ 端多聚A不是由 DNA 转录的,而是转录后再加上的。细胞核 DNA 中并不含有大于 50 个核苷酸的多聚 (dA-dT) 顺序,而 hnRNA 与 mRNA 却含有直至 200 个核苷酸的多聚A片段。

3′-脱氧腺苷能抑制 rRNA 前体合成和 mRNA 在细胞质中出现,但不影响 hnRNA 合成。进一步研究表明 3′ 脱氧腺苷能抑制多聚A合成,结果阻止多聚A加到 hnRNA 与 mRNA 分子上。由于多聚A的添加是 mRNA 成熟的重要一步,也是 mRNA 进入细胞质所必须的,所以上述结果并不影响 hnRNA 与 mRNA 是前体与产物关系。细胞核或细胞质中 RNA 加上多聚A的动力学是很复杂的,不能由此得出 hnRNA 与 mRNA 之间存在着简单的前体与产物关系的结论。细胞核中多聚A顺序的消失,一部分是由于多聚A随 mRNA 进入细胞质,也有一部分是在细胞核内被降解的。另外在细胞质中也存在着加上多聚A的作用。

(5) hnRNA 与 mRNA 中存在着相同的顺序。若 hnRNA 是细胞质 mRNA 前体，那么 hnRNA 将含有存在于细胞质 mRNA 中的所有顺序。不同细胞 hnRNA 和细胞质 mRNA 与细胞核 DNA 进行竞争杂交的实验清楚表明,在细胞质 mRNA 中存在的顺序都在 hnRNA 中，但在 hnRNA 中却存在着许多细胞质 mRNA 中不存在的顺序。这一事实虽不是最后证明,但却有力地支持 hnRNA 是细胞质 mRNA 前体的想法,同时也表明 hnRNA 的顺序要比细胞质 mRNA 复杂得多。 hnRNA 中大部分顺序是不进入细胞质的。目前对留在细胞核内的顺序知道得还很少，至于为何选择这些顺序留在细胞核内则更不清楚。

大多数 mRNA 是从非重复顺序 DNA 转录而来。hnRNA 中重复顺序要比 mRNA 多,这表明绝大部分重复顺序已在细胞核内降解。hnRNA 中重复顺序大多在 5′ 端,而非重复顺序在 3′ 端,因此很可能 mRNA 主要从邻近 hnRNA3′ 端的非重复顺序得来。(图5-22)。

(6) 在 hnRNA 中存在特异性 mRNA 顺序。hnRNA 能和与细胞质特异性 mRNA 互补的 DNA 或 RNA 进行杂交。与鸭珠蛋白 mRNA 互补的 DNA 能与鸭红细胞 28-100S hnRNA 进行杂交，这表明在较大的鸭红细胞 hnRNA 中存在有珠蛋白 mRNA 顺序。卵清蛋白 mRNA 顺序亦在鸡输卵管 hnRNA 中测到。

在病毒转化细胞中，hnRNA 能与转化细胞所用的病毒 DNA 或 RNA 进行杂交。SV40 转化的细胞中含有 SV40DNA，它与细胞染色体共价结合。 通过杂交试验可以证明在转化细胞 hnRNA 与多核糖体 RNA（主要是细胞质 mRNA）中存在 SV40 特异性顺序。双股 SV40DNA 分子量为 3×10^6d,而含有 SV40 特异性顺序的 hnRNA 分子量为 1.5—4×10^6d,含有 SV40 特异性顺序的 mRNA 分子量为 0.5×10^6d。这表明 SV40 转化细胞转录的 RNA 要比单股 SV40DNA 或细胞质中发现的 SV40 病毒特异性 mRNA 要长。另外在腺病毒与疱疹病毒转化的细胞中都已发现存在着病毒特异性顺序的 hnRNA 与细胞质 mRNA。鼠乳腺肿瘤病毒特异性 RNA 顺序亦存在于肿瘤细胞 hnRNA 与细胞质 mRNA 中。所以在 hnRNA 中存在着与正常顺序共价连结的病毒特异性顺序，这可能是病毒转化细胞的一种普遍现象。这一点尽管还不能最终证明 hnRNA 是 mRNA 前体,但与 hnRNA 是 mRNA 前体的想法相符。

(7) hnRNA 能被翻译为特定蛋白质。hnRNA 能为核糖体特异性识别,表明 hnRNA 含有与核糖体特异性结合位置。这种结合类似于 mRNA 与核糖体结合，需要特异性蛋白因子(可能就是起动因子),同时这种结合又为蛋白质合成起动抑制剂所抑制。

hnRNA 中存在 mRNA 顺序的更为直接的证据是 hnRNA 能在体外蛋白质合成系统中被翻译。如合成珠蛋白细胞(鼠胚胎肝细胞与鸡非成熟红细胞等)的 hnRNA,在爪蟾卵母细胞中与无细胞蛋白质合成系统中都能促进珠蛋白合成,而用脑细胞 hnRNA 就观察不到珠蛋白的合成。来自各种组织的 hnRNA 都可在体外蛋白质合成系统中被翻译为多肽,但翻译产物往往要比 mRNA 复杂。 如鸡非成熟红细胞 hnRNA 在无细胞蛋白质合成系统中翻译为珠蛋白的量不到总翻译量的 20%,而其细胞质 mRNA 翻译产物 70% 以上是珠蛋白。如用鸡网织红细胞 hnRNA 进行体外翻译，所产生的珠蛋白 α 与 β 链之间比例亦与其珠蛋白 mRNA 翻译产物不同。 此种差别不仅是因为 hnRNA 种类复杂，还可能与 hnRNA 有比 mRNA 更为复杂的二级与三级结构有关。

综上所述，hnRNA 与 mRNA 之间具有密切关系，虽然不能说全部，至少部分 hnRNA 是 mRNA 的前体。实验表明，只有 10—40% hnRNA 分子具有多聚 A，而 3′ 端的多聚 A 被认为是 mRNA 进入细胞质所必须的，那么不具有 3′ 端多聚 A 的 hnRNA 分子是不是 mRNA 前体？没有多聚 A 的被严格限制在核内的 hnRNA 功能是什么？这些问题目前还不能得到满意的回答。

在分离纯化 hnRNA 时遇到一个很大的困难，就是 hnRNA 很易聚合，在温和条件下 hnRNA 解聚不完全，而在强烈变性条件下则又会引起 hnRNA 降解，因此要分离纯化到完整的 hnRNA 并准确地测定其大小是困难的。

hnRNA 的分子结构

上节提到，hnRNA 与 mRNA 具有相似的 5′ 帽子结构和 3′ 多聚 A 末端。由于发现有形成帽子结构和附加多聚 A 的酶，(见“mRNA 前体的加工”一节)，因此，hnRNA 并不是转录的原始产物。此外，hnRNA 的 5′端一般有 3 个寡聚 U 短片段，每个长约 30 个核苷酸，这些片段似由 DNA 转录而来。图 5-22 是 hnRNA 的分子结构示意图。



图 5-22　hnRNA 的分子结构示意图

hnRNA 分子中分散着重复顺序与非重复顺序，它们的数量、长短和分布随细胞种类不同而不同。鼠腹水细胞 hnRNA 中，28% 是 1,200 个核苷酸长的非重复顺序分散在 200 个核苷酸长的重复顺序中，而 43% 则是由 900 个核苷酸长的非重复顺序分散于 200 个核苷酸长的重复顺序中。HeLa 细胞 hnRNA3′ 端多聚 A 旁有一非重复顺序，它的长度类似或稍大于其 mRNA 平均长度。除这一特殊区域外，HeLa 细胞 hnRNA 中非重复顺序均匀地分散于重复顺序之中。海胆胚胎 hnRNA 中也存在大量 800 个核苷酸长的非重复顺序，它们分散于 300 个核苷酸长的重复顺序之中。hnRNA 的主要部分是由 DNA 中非重复顺序和重复顺序转录而来。

hnRNA 的另一特点是大多数 hnRNA 都含有茎环结构即 dsRNA 区，这种茎环结构是由分子内碱基配对形成的。这种茎环结构并不多，可能位于或靠近编码蛋白质的非重复顺序两端。将 hnRNA 用核酸酶处理而后用羟基磷灰石柱或纤维素柱吸附可分离制备到茎环区。鼠肝细胞 hnRNA 中茎环区长约 60 个核苷酸。鼠腹水细胞 hnRNA 中茎环区长分别为 10—20 个与 100 个核苷酸。

和细胞质 mRNA 一样，在 hnRNA 中亦已发现 N^6-甲基腺苷存在。

真核生物 mRNA 基因中的插入顺序(内含子)被转录在 hnRNA 分子中

在 DNA 结构一章曾经提到,许多真核蛋白质基因含有内含子。这种结构基因的不连续现象正在越来越多地出现,看来这是真核生物所特有的普遍现象。除了第二章提到的例子外,下面二个基因也是研究较为详细的。鸡卵白蛋白基因含有多至七个内含子和八个外显子(图 5-23)。腺病毒 Ad2 的 Hexon (一种主要外壳蛋白) mRNA 基因在外显子间存在三个内含子,见图 5-24。

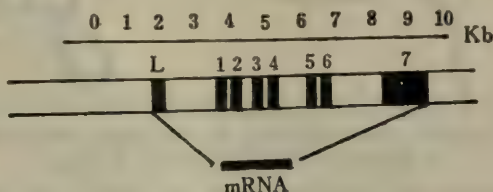


图 5-23 鸡卵白蛋白基因中的插入顺序

部分黑色为外显子;外显子间为内含子。

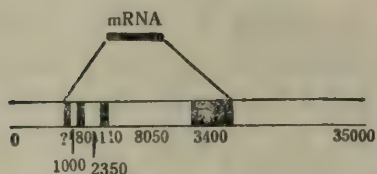


图 5-24 Ad2 中 Hexon mRNA 基因区域

黑色部分代表外显子;外显子间为内含子;

图中数目字代表核苷酸对数。

从已研究的蛋白基因中内含子情况可以看出,在一个基因中往往有多个内含子,同时,内含子的总链长比结构基因编码部分还要长,甚至长得多。至于内含子的生物功能,目前还是很不清楚的,除了第二章提到的可能功能外,近年来有的实验指出,个别基因中的内含子是另一基因的外显子,表明这段 DNA 序列具有明确的生物功能。还发现,在酵母线粒体的个别内含子中同时存在起始密码子和终止密码子,它们编码一个 235 个氨基酸的蛋白质,表明内含子本身可能是一个独立的结构基因。

由于成熟 mRNA 并不含有这些内含子,因此有人曾经提出,可能在 RNA 聚合酶进行转录时,酶能够"跳过"形成环形结构的插入顺序,即内含子不被转录。但不久这种假设被否定,而代之以下述结论:内含子和外显子一起被转录为原始转录产物,然后加工为 mRNA。

hnRNA 含有与内含子互补顺序是内含子与外显子一起被转录的很有力的证据。β 珠蛋白 mRNA 前体能与 β 珠蛋白基因完全互补,实验是这样进行的:用同位素短时间标记细胞核 RNA,得到一个 15S RNA,它具有与 10S 的 β 珠蛋白 mRNA 相似的 5′ 端帽子结构和 3′ 端多聚 A 顺序,而且与 β 珠蛋白基因的长度相当,因此,一般认为 15S RNA 是 10Sβ 珠蛋白 mRNA 的前体。将 β 珠蛋白基因片段分别与纯化的 15S RNA 和 10S mRNA 杂交,在电镜下观察形成的 R 环 (RNA 与 DNA 杂交后所形成的环),结果发现 15SRNA

与 DNA 形成的R环是完全互补的，而10SmRNA 与 DNA 杂交时，则有 DNA 环出现，说明15SRNA 中含有与内含子互补的顺序，在10SmRNA 中，这个顺序已被删去(图5-25)。

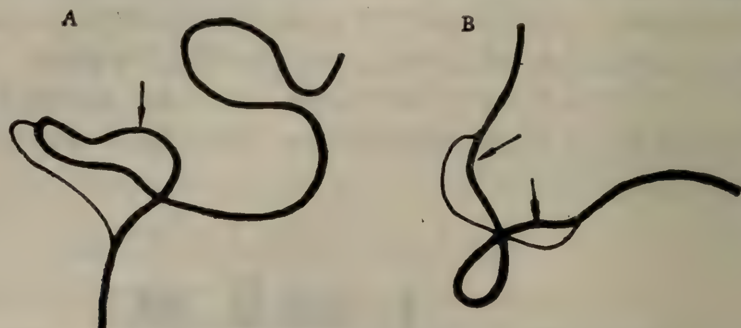


图 5-25 15S 珠蛋白 mRNA 前体 (A) 和 10S 珠蛋白 mRNA(B) 分别与含珠蛋白基因的 DNA 片段杂交的电镜示意图

箭号指出的R环是 RNA-DNA 杂交链，粗线为双链，细线为单链。

用 S_1 核酸酶降解技术也说明了免疫球蛋白轻链mRNA 前体上含有与内含子互补的顺序。将轻链 mRNA 在引物存在下进行反转录，可以得到两种产物，一种在不变区 (C) 内终止，cDNA 链长为200—300核苷酸；另一种含有不变区和可变区 (C + V)，其链长约500—700核苷酸。用同位素标记 cDNA，然后与 DNA 或 RNA 杂交。如果这些 DNA 与 RNA 是连续的(即无内含子)，那么杂交体是完全抗 S_1 核酸酶的；如果含有内含子的 DNA 或 RNA 与之杂交，由于 S_1 酶是专一水解单链或错配对区的(见第三章)，因此将得到的是部分抗 S_1 核酸酶的结果。用 ^{32}P 标记的 (C + V) cDNA 与纯化的轻链 mRNA 杂交后用 S_1 酶处理，得到约700核苷酸的标记 DNA，说明是完全抗 S_1 酶的；若用轻链 mRNA 的27S前体进行同样杂交，酶解后，仅得到 290 核苷酸长的标记 cDNA，表明是部分抗 S_1 酶的；用含轻链基因的 DNA 片段与 (V + C)cDNA 杂交，经 S_1 酶处理，也得到295核苷酸长的抗 S_1 酶的 cDNA。上述实验充分说明，在免疫球蛋白轻链 mRNA 前体上含有在可变区与不变区间的内含子，同时该顺序是与结构基因同时转录的。

目前尚没发现内含子不被转录的例子，rRNA、tRNA 基因中的插入顺序(如果存在的话)也都是同时转录在前体分子之中的(见下文)。

真核生物 mRNA 基因结构及上游的转录调控部位如图5-26所示(也参阅图5-2)，原始转录产物中除了内含子外，也包括两端的非编码区。

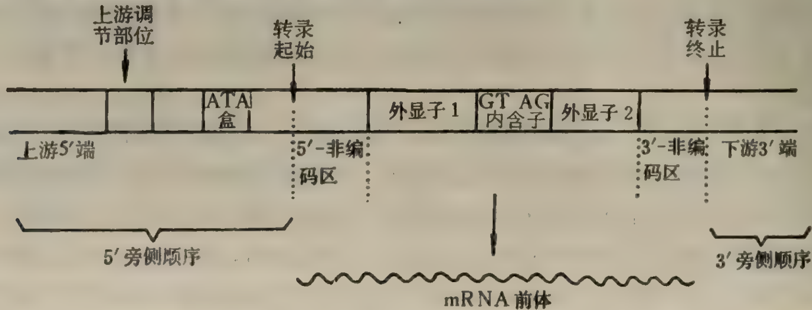


图 5-26 真核生物 mRNA 基因结构及上游调控部位

mRNA 前体的加工

原始转录产物通过 hnRNA 变为成熟的 mRNA，需要经过下列一些步骤:

1. hnRNA 被剪接 (splicing)，除去由内含子转录来的序列，将含信息的结构部分联接起来；

2. 5′ 端帽子的形成；

3. 3′ 端多聚 A 的附加；

4. 内部顺序的甲基化。

关于 mRNA 前体的剪接

在 mRNA 的原始转录产物中，内含子转录顺序是如何被剪接的呢？目前，除了个别报道如四膜虫 rRNA 前体的剪接可以自我催化外(见下文)，一般认为是存在一个剪接酶系，但是迄今并不像 rRNA 和 tRNA 加工酶那样已被纯化；同时由于加工可能在核蛋白颗粒 (hnRNP) 内进行，所以情况更为复杂。后来发现有些由细胞核小分子 RNA 构成的核蛋白 (snRNP) 离心时常与 hnRNP 一起沉降，因而提出 snRNA 可能与 hnRNA 剪接有关。关于 snRNA 的结构与功能见第八章。

剪接酶和 snRNA 等究竟怎样与 hnRNA 发生关系呢？比较了近四十个内含子与外显子联接部位附近的核苷酸序列，并没有找到令人信服的通常为酶识别的二元对称结构，但是在接头处都存在不太长的同源顺序(见图 5-27)。这个同源顺序正好与某些 snRNA 如 UlRNA 有约 15 对碱基配对，因此推测正是形成这样的特殊高级结构而为酶所剪接。关于 UlRNA 与 hnRNA 中内含子和外显子联接部位互补可能方式见第八章图 8-2。

	外显子			内含子																外显子
C	17	1	5	1	0	2	3	4	5…9	15	14	8	15	18	17	8	29	0	0	7
A	17	31	4	1	0	25	27	4	5…0	3	3	9	7	1	0	10	2	42	0	4
G	2	2	30	37	0	8	4	33	3…3	1	4	2	1	0	1	8	0	1	43	23
U	3	5	0	0	39	4	5	1	26…31	23	22	24	19	24	25	17	12	0	0	
	A	A	G	G	U	A	A	G	U…U	U	U	U	U	U	U	U	C	A	G	G
	C									C	C		C	C	C	C				

外显子 ←—— 内含子 ——→ 外显子

图 5-27 mRNA 前体中内含子与外显子联接部位附近的核苷酸的分布几率用数字表示及其同源性

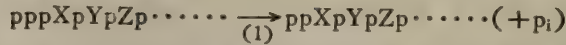
关于 mRNA 剪接酶的纯化及其作用方式目前了解得还很少。

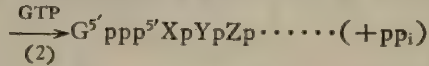
5′ 端帽子的形成

5′端帽子结构存在于 hnRNA 中，它在转录的早期阶段甚至可能在转录终止前形成。帽子结构的形成与 3′ 端附加多聚 A 和内含子的剪切都可能与核中 mRNA 前体转变为细胞质 mRNA 有关。关于戴帽机制主要来自对病毒 mRNA 的研究结果，实验表明戴帽有两种途径:

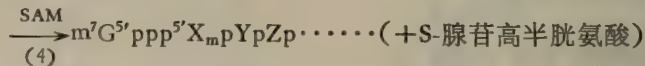
(1) 转录起始不久进行戴帽——从原始转录产物 5′ 端三磷酸脱去一个磷酸，然后与

GTP 反应生成 5′,5′ 三磷酸联结的键，并释放出无机焦磷酸；最后以 S-腺苷甲硫氨酸(SAM) 进行甲基化产生帽子结构。





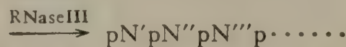


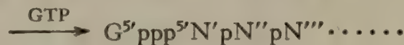


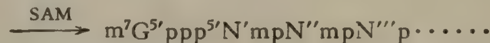
以这种方式戴帽的为呼肠孤病毒和牛痘病毒 RNA。催化上述反应的酶分别是:(1)RNA三磷酸酶，(2) mRNA 鸟苷酰转移酶,(3) mRNA (鸟嘌呤—7) 甲基转移酶，(4) mRNA (核苷-2′)甲基转移酶,也叫 2′-*o* 甲基转移酶。这些酶都已从牛痘病毒纯化，从其他病毒和动物组织中也测到或纯化了这些酶。牛痘病毒中的酶 (2) 和酶 (3) 无论用亲和层析或离子交换柱层析都很难分开,纯化的酶(2)分子量为 127,000d,含二条链(分子量分别为 95,000d 和 31,400d)。此酶专一底物是 5′ 端为二磷酸的 mRNA 或多聚 A 或多聚 G。酶(3)的甲基供体是 SAM,专一甲基化帽子上的 G。2′-o-甲基转移酶也是以 SAM 为供体，专一甲基化 mRNA5′ 端的第一个核苷酸的核糖 2′-羟基。酶分子量为 38,000d，以 pN、ppN 或 $G^{5'}$ $^{5'}$pppN 结尾的 RNA 不能作为酶的底物,而末端为 $m^7G^{5'}ppp^{5'}N$ 的多聚 A 和多聚 I 是最好的底物。

(2) 戴帽是在剪切后进行——这种戴帽方式以疱疹口炎病毒 RNA 为代表，首先用剪切酶(可能是 RNaseIII)切开转录的 RNA 链，产生 5′ 磷酸末端，然后形成 5′,5′-三磷酸联结的键，并甲基化得到帽子结构。









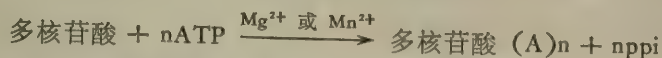
甲基转移酶抑制剂可用于研究甲基化对 mRNA 合成、输送及在翻译中的影响。S-tubercidinyl homocysteine 能有效地抑制 Novikoff 肝癌 mRNA 帽子上的核糖 2′-甲基化和 RNA 内部的 A 的 6-甲基化,但不影响帽子 m^7G 的生成。由于该抑制剂并不抑制细胞的生存,所以核糖的 2′-甲基化或链内部 A 的 6-甲基化被认为对于 mRNA 从细胞核输送至细胞质以及随后进行翻译是不必需的。环亮氨酸 (Cycloleucine) 是 L-Met S-腺苷基转移酶的竞争性抑制剂，能阻断肉瘤病毒 RNA 帽子上的核糖和链内部的甲基化，但不影响帽子上 G 的 7-甲基化，对病毒复制也无明显影响,表明核糖和内部甲基化对于该病毒 RNA 的合成,加工和输送也不是必需的。

相反，腺苷高半胱氨酸类似物——5′-脱氧-5′-异丁酰基腺苷却能抑制劳氏肉瘤生长 90% 以上，分析表明，该抑制剂可能抑制 mRNA (鸟嘌呤-7) 甲基转移酶，而这个酶对

mRNA 合成是很重要的。

3′端多聚A的附加

多聚A的附加是在多聚A聚合酶即 RNA 末端腺苷酸转移酶催化下完成的，该酶以具 3′OH 的 RNA 为受体，ATP 作供体，需 Mg^{2+} 或 Mn^{2+}。其他 NTP 特别是 GTP 是聚合酶的抑制剂。大部分多聚A聚合酶存在于细胞核内，但在某些来源的细胞质中也发现有酶的活性。



$$\text{多核苷酸} + n\text{ATP} \xrightarrow{Mg^{2+}\ \text{或}\ Mn^{2+}} \text{多核苷酸}\ (\text{A})n + n\text{ppi}$$

对腺病毒 mRNA 的研究揭示，转录时 RNA 聚合酶并不在 mRNA 末端处停止，因此，在附加多聚A前需要修剪 3′末端。前面提到，hnRNA 和成熟 mRNA 两者都能测出多聚A尾巴。随着生活时期的变化，多聚A的链长逐渐缩短。

序列分析表明，在 mRNA-多聚A接头附近，并无同源顺序。但在许多真核 mRNA 的多聚A 5′端起约 13—20 核苷酸处，发现一个同源区 AAUAAA，这被认为可能是多聚A附加的识别部位。

mRNA 的内部甲基化

mRNA 内部甲基化主要指 m^6A 的生成。一些实验说明，含 m^6A 的顺序在 hnRNA 中已经存在，同时在加工中被保存。一般来说，m^6A 在真核生物或病毒 mRNA 的内部位置是不确切的，但在 HeLa 和L细胞中，在 SV40 晚期 mRNA 中分离到两个含 m^6A 顺序：Apm^6ApCp 和 Gpm^6ApCp。由于许多病毒和细胞 mRNA 并不存在 m^6A，似乎这个修饰成分对 mRNA 翻译不是必要的。

我们已经讨论了 DNA 模板、RNA 的原始转录产物和 mRNA 成熟的各个步骤。这些过程大多是在细胞核内完成的。但是 mRNA 在细胞内的状态是怎样的呢？它是怎样由细胞核运到细胞质中的呢？现在了解到在真核生物中，mRNA 在其生命的所有过程中都是以核蛋白复合物形式存在的，这种复合形式可能是一种调节 mRNA 由细胞核向细胞质运输的形式，也可能是细胞内保卫 mRNA 的一种方式，还可能是在 mRNA 的翻译水平上调节蛋白质生物合成的一种形式。这种 mRNA 与蛋白质组成的复合物不同于核糖体（ribosome）（见第六章），曾被称为信息体(informasome)。信息体的沉降系数与其中 RNA 的沉降系数有关，一般比 RNA 大 2—2.5 倍；当 mRNA 分子不均一时，信息体显示出不均一的沉降分布，但如仅一种 mRNA 存在，信息体是均一的。信息体在 CsCl 中的沉降密度为 $1.46g/cm^3$ 到 $1.36g/cm^3$，介乎蛋白质（1.2—1.3 g/cm^3）和 RNA(1.9—2.0g/cm^3)之间，用放射性尿嘧啶或放射性氨基酸标记时，信息体都带有放射性，说明信息体含有 RNA 和蛋白质。

实验证明信息体不是人为的矫作物，其中的 RNA 是 mRNA（能与 DNA 杂交，在无细胞系统中被翻译），从不同来源的信息体中可分离出多种多肽或蛋白质，其中许多组分的分子量已经测定。

hnRNA 也是以核蛋白颗粒形式存在，即生成 hnRNA 颗粒，在这些颗粒中约含 20% RNA 和 80% 蛋白质。有人已经测出 hnRNA 颗粒中存在多聚A聚合酶、戴帽的酶、核

酸酶和蛋白激酶等酶的活性，甚至认为 hnRNA 剪接酶系也可能存在于 hnRNP 颗粒中，这对于 hnRNA 加工的研究无疑是很有意义的。

五、rRNA 的转录与加工

rRNA 的结构与功能分别在第三章和第六章进行讨论。以 rDNA 为模板在 RNA 聚合酶（真核生物为 RNA 聚合酶 I）催化下得到的原始转录产物是 rRNA 前体，然后经过逐步加工成为成熟的 rRNA 分子。

rRNA 基因（rDNA）与 rRNA 前体

在细菌基因组中，一个转录单位内含三个 rRNA 基因，同时插入多个 tRNA 基因按一定顺序排列在一起。构成大肠杆菌的转录单位约有 5,300—5,500 碱基对，转录后得到 30S 的 rRNA 前体。前体分子可从缺乏 RNaseIII 加工酶的突变株中分离。

真核 rRNA 基因一般是中度重复顺序，不同生物的拷贝数在五十至几千之间，它们成簇排列在一起。哺乳动物的每一重复单位包括 18S、5.8S 和 28SrRNA 序列，彼此被隔间区所分开，转录产生一个长的 rRNA 前体。不同生物的 rRNA 前体大小是不一样的，如哺乳动物为 45S，爪蟾 40S，果蝇 38S，酵母37S。rDNA 中插入顺序（IVS）存在情况也有不同，有的不存在 IVS，有的有一个或两个 IVS；个别基因内带有缺口（gap）；线粒

表 5-8 rRNA 基因（rDNA）的种类

生物来源	部位	基因顺序	rRNA 前体
原核生物	染色体	16S-tRNA-23S-5S-rRNA	30S
大鼠、小鼠、人	核仁	18S-5.8S-28S	45S
爪蟾	核仁	18S-5.8S-28S	40S
果蝇	染色体	18S-5.8S-2S-$28S_{\alpha}$-缺口-$28S_{\beta}$	38S
酵母	外染色体	17S-5.8S-26S	37S
绒泡菌	外染色体	19S-5.8S-$28S_a$-IVS_a-$28S_b$-IVS_b-$28S_c$	40S
四膜虫	外染色体	17S-5.8S-$26S_a$-IVS_a-$28S_b$	>35S
利什曼虫	外染色体	16S-5.8S-$26S_{\alpha}$-缺口-$26S_{\beta}$	
小鼠、人	线粒体	tRNA-12S-tRNA-16S-tRNA	
酵母	线粒体	15S-tRNA-$21S_a$-IVS-$21S_b$	
链孢霉	线粒体	17S-tRNA-$25S_a$-IVS-$25S_b$-tRNA	
玉米、菠菜	叶绿体	16S-tRNA-23S-4.5S-5S	
眼虫	叶绿体	16S-tRNA-23S-5S	
衣藻	叶绿体	16S-tRNA-7S-3S-$23S_a$-IVS-$23S_b$-5S	

* IVS 代表插入顺序。
摘自 *Ann. Rev. Biochem.*, 1982, **51**, 642。

体和叶绿体的 rDNA 具有与原核生物类似的基因排列，但本身存在某些特点（见第八章）。表 5-8 列出已知 rRNA 基因的种类。从表 5-8 可以看到 rDNA 具有下列一些特性：

(1) 基因的极性是保守的：即 16—18S RNA 在前体的 5′ 端，而 23—28S RNA 在3′ 端。

(2) 原核生物和叶绿体基因的 3′ 端含 5S rDNA。

(3) 原核生物、叶绿体和线粒体的 rRNA 基因中连有 tRNA 基因，有的在内部，有的在两端。

(4) 真核生物 rRNA 基因与 5S rDNA 和 tDNA 分离，代之以 5.8S rDNA，并嵌在 16—18S rDNA 与 26—28S rDNA 之间。不同真核生物 5S rDNA 虽都是一个独立的转录单位，但分布却有不同（见第二章）。

(5) 高等植物叶绿体基因含 4.5S rDNA。

rDNA 经过转录得到与 rDNA 同样大小的 rRNA 前体，其中包括 tRNA 前体、间隔区以及插入顺序等。在正常情况下，迅速被加工为成熟的 rRNA 和 tRNA 分子。但是某些缺乏加工酶的变异株或用氯霉素处理细胞使蛋白质合成受阻，则大量堆积核酸原始前体或中间前体，这是研究加工过程和加工酶性质的很好材料，许多结果就是通过分离鉴定中间前体而得到的。

rRNA 前体的加工

rRNA 前体的加工也是在一系列酶催化下进行的。与 rRNA 前体加工有关的酶主要有二类：(i) 能使前体被剪切成一定链长的 rRNA 分子，这类 RNA 加工酶中研究得较多的有大肠杆菌 RNaseIII，RNaseE，枯草杆菌 RNaseM5 等。真核生物也有类似的加工酶活性。(ii)修饰酶，催化 rRNA 中修饰成分的形成。rRNA（除 5S RNA 外）都含有修饰成分，主要是核糖 2′ 位的甲基化以及一些特殊的修饰碱基（如 $m^1ncp^3\psi$ 等见第一章）。对 rRNA 修饰酶的研究远没有 tRNA 修饰酶的详细和深入。

图 5-28 是大肠杆菌 RNA 加工变异的基因图，RNA 加工基因在圆圈的内部，其中与 rRNA 有关的是 rnc(RNaseIII 基因)和 rne(RNaseE 基因)，其他几个则与 tRNA 加工有关（见第六节）。由于原核生物的 rDNA 中通常插有 tDNA，因此其原始转录产物的

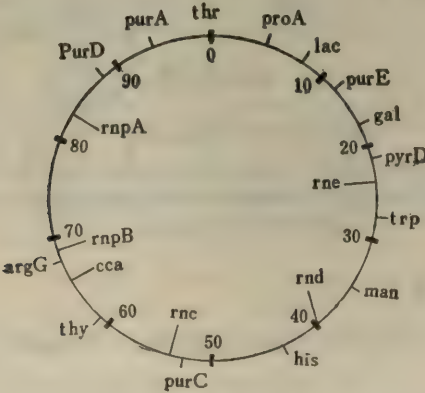


图 5-28 大肠杆菌 K-12 中 RNA 加工变异的基因图

rnc, RNaseIII; rne, RNaseE; rnd, RNaseD; rnp, RNaseP (rnpA 变异影响 RNaseP 中的多肽组分，rnpB 变异则影响其中的 RNA 组分); cca, tRNA 核苷酰转移酶。

加工是两者同时进行的。

RNaseIII 是一个内切核酸酶，虽然在一定条件下能作用于单链 RNA，但该酶的识别部位为特定的螺旋区。1973 年发现它与 rRNA 前体加工有关，它水解 rRNA 前体产生 16S 和 23S rRNA 部分。RNaseIII 由两个相同的分子量为 25,000 的亚基构成，大概不含 RNA 组分(用于 tRNA 前体加工的 RNaseP 含 2/3 RNA，见第六节)。RNaseIII 识别 rRNA 前体以及 T7 噬菌体早期 mRNA 前体的位点已经确定，被认为与 RNA 双螺旋区的三级结构有关。RNaseE 作用于较大前体产生 5S RNA 前体(比成熟 5S RNA 在两端各长三个核苷酸)，天然酶的分子量为 70,000d，亚基结构不详。RNaseE 的识别位点也是螺旋区域。RNaseM5 加工枯草杆菌 5S RNA 前体得到成熟的 5S RNA，裂解点亦为双链区。部分纯化的酶由 α 与 β 两个亚基组成，β 亚基具底物结合能力，而 α 亚基为催化亚基。大肠杆菌中使 16S 和 23S rRNA 成熟的酶分别叫 RNase M16 和 RNase M23，它们都是内切核酸酶。

某些真核生物(如四膜虫、绒泡菌) rRNA 前体含插入顺序，通过剪接切除这些插入顺序可能和 mRNA 前体加工一样，也是在一个酶系催化下完成的。但是 1982 年 Cech 研究组发现，四膜虫 rRNA 前体含 413 个核苷酸残基的插入顺序，可以在完全不存在蛋白质(酶)条件下自我准确地切除，表明 RNA 分子本身具有催化活性(图 5-29)，这是一个很有意义的发现。

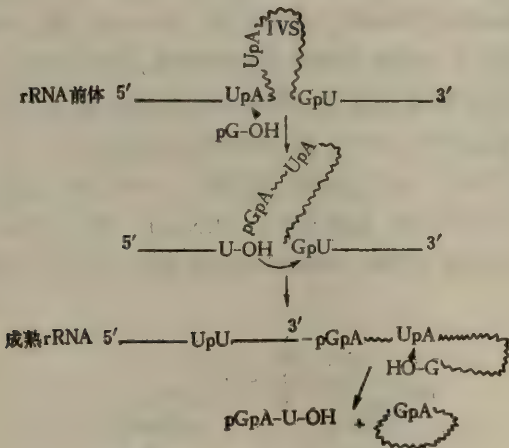


图 5-29 RNA 的自我剪接

一个鸟苷辅助因子在插入顺序 5′ 端通过亲核反应启动一系列反应而切除插入顺序，同时把外显子联结起来。图中也表示被切除的插入顺序产生第二次亲核反应得到特定的最终产物。

原核生物 rRNA 的加工

大肠杆菌 rRNA 基因的原始转录产物沉降常数为 30S(p30)，分子量 2.1×10^6d，约含 6,300 核苷酸，5′末端为 pppA。通过分子杂交、中间前体分离及其末端鉴定、酶解图谱分析等技术，已基本弄清大肠杆菌 rRNA 前体的加工过程。表 5-9 列出 rRNA 前体与成熟 rRNA 某些性质的比较。从表 5-9 可以看出：(1) p30 中有 80% 左右的核苷酸顺序被加工为成熟 rRNA；(2) p16 与 m16 及 p23 与 m23 分别相差 250 或 500 核苷酸

长度，其 5′ 端和 3′ 端分别为不同的核苷酸顺序，说明要在两端同时剪切或修剪才能得到成熟产物。也已分离出更接近于 16S rRNA 的 16.3S 中间前体，比 16S 只多 60 个核苷酸。(3)分离出三种 5S rRNA 中间前体分别只在 5′ 端多三个，二个和一个核苷酸，而 3′ 端完全相同。在变异株中也找到了更大的 9S 中间前体，它被 RNaseE 水解生成 5S rRNA。

表 5-9 大肠杆菌 rRNA 前体与成熟 rRNA*

rRNA	分子量	核苷酸残基	5′末端	3′末端
p30rRNA	2.1×10^6	~6300	{pppACUG pppAUCG	
p16rRNA	0.6×10^6	~1600	pUG……	……$CACACA_{OH}$
m16rRNA	0.55×10^6	1542	pAAAUG……	……$CUUCA_{OH}$
p23rRNA	1.15×10^6	~3400		
m23rRNA	1.1×10^6	2904	pGGU……	
p5rRNA-1	~0.04×10^6	123	pAUUUG……	……$GGCAU_{OH}$
p5rRNA-2		122	pUUUG……	$GGCAU_{OH}$
p5rRNA-3		121	pUUG……	$GGCAU_{OH}$
m5rRNA		120	pUG……	$GGCAU_{OH}$

* p 表示前体，m 代表成熟 rRNA。大肠杆菌全部三种成熟 rRNA 的一级结构都已测出(见第三章)。

大肠杆菌 p30 的半寿期只 2 分钟，故在正常细胞中迅速被加工为成熟 rRNA。RNaseIII 缺陷型可使 p30 堆积，当用 RNaseIII 体外作用于 p30，并分离出其中间产物时，发现 p16(16SrRNA 的中间前体)的 5′ 端与 p30 相同，均为 pppA，故确定在 p30 中 p16 位于其 5′ 端。p30 的加工过程见图 5-30，图中也包括 tRNA 的加工(关于 tRNA 加工酶见第六节)。

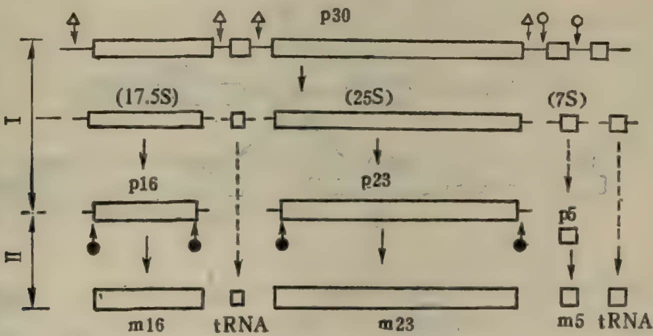


图 5-30 大肠杆菌 rRNA 前体加工过程

I 表示第一阶段加工，⇕为 RNaseIII 切点，♀是 RNaseE 切点；II 表示第二阶段加工，↕为 RNaseM 切点；tRNA 加工酶是 RNaseP 和 RNaseF，同时进行剪切。

在不存在蛋白质生物合成条件下，可分离出两个枯草杆菌 5S RNA 前体：p5A 与 p5B，前者为 179 核苷酸，后者 140—150 核苷酸，它们都比成熟 5S RNA (118 核苷酸) 长数十个核苷酸之多。由于 p5A 与 p5B 的核苷酸序列有差，因此它们可能为不同基因

的转录产物。 p5A 作为 RNase M5 体外底物，要求有紧密的螺旋结构；同时实验指出，p5A 中两端冗长的序列对 RNase M5 与 p5A 相互作用关系密切。

原核生物 rRNA 含有较多的甲基化修饰成分，特别是 2-甲基核糖。比较成熟 rRNA 及其前体中的甲基化成分可看出，p30 含有 23S rRNA 中的全部甲基化成分，但 16S rRNA 中修饰成分较少，这表明在 rRNA 前体中两个 RNA 组分的修饰不是协同的，即 23 S rRNA 的修饰在加工的早期阶段，而 16S rRNA 在后阶段。

原核 16S rRNA 和 23S rRNA 中修饰成分见第二章，其中 m^4Gm 看来是 16S rRNA 一个特征成分。一般 5S RNA 中无修饰成分，故不存在修饰加工问题。

真核生物 rRNA 的加工

真核生物 rRNA 的前体具有与原核生物不同的结构特征。 通常真核生物有更大的前体，见表 5-8，最大的哺乳动物的原始转录产物为 45S(4.6×10^6d)，较低等真核生物也有 36—38S($2.6—2.8 \times 10^6$d)。其次，真核生物 5S RNA 前体由独立于其他三种 rRNA 的基因转录；线粒体中未发现 5S RNA，而叶绿体的 5S RNA 则与原核生物一样存在于三种 rRNA 组成的前体之中(见表 5-8 和第八章)。另外，在真核的 rRNA 前体中含有独特的 5.8S rRNA，它位于其他二个大的 rRNA 之间并为间隔区所隔开，由于 5.8S rRNA 与 28S rRNA 间具有较长的互补氢键对，因而只在变性后才释放出 5.8S rRNA。

从原始转录产物加工产生成熟 rRNA，真核生物与原核生物相比，过程更为复杂，同时也可能存在不同的加工方式。将正在生长的哺乳动物细胞置于 ^{3}H 或 ^{14}C 尿苷中培养短时间(5—10 分钟)，则 45S RNA 迅速被标记，延长标记时间，45S RNA 放射性增加，20 分钟后，放射性出现在 32S 和 18S RNA 中，更长时间后(约60 分钟)28S RNA 被标记。这

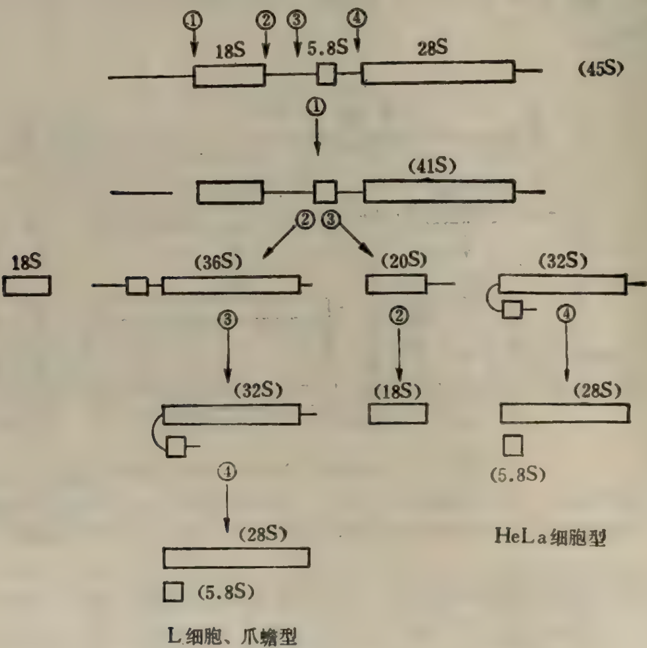


图 5-31　真核生物 rRNA 前体的二种加工方式

表明 rRNA 前体加工是以特定方式进行的。用 ^{3}H-尿苷脉冲标记后给与放线菌素 D，由于阻断了 45S RNA 的合成，因而使 45S RNA 消失，而增加了 32S RNA 和 18S RNA 组分，以及稍后使 28S RNA 中的放射性堆积。45S RNA 前体转变为 28S RNA 和 18S RNA 是一个多步加工过程，用聚丙烯酰胺凝胶电泳可从细胞核 RNA 中分离出许多中间前体，沉降常数有 41S、36S、24S 和 20S 等，标记动力学也证实它们是 rRNA 生成途经中的真正中间物。结合其它分析数据，提出了 rRNA 前体可能存在二种加工类型(见图 5-31)，一类是 18S 与 28S rRNA 近乎同时成熟(HeLa 细胞型)，另一类则是 18S 先于 28S RNA 出现(L 细胞、爪蟾型)。图中也表示了 5.8S rRNA 的加工。

一些真核生物 rRNA 的加工过程列于表 5-10。由于中间前体的不断发现，还可能有补充和修改。

表 5-10 不同真核 rRNA 前体加工过程的比较

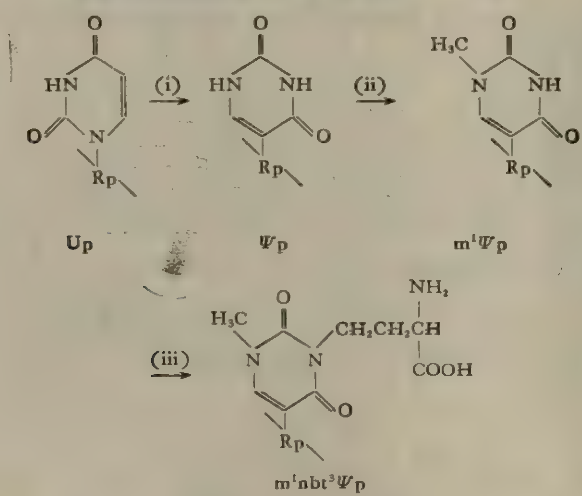
来源	原始转录产物	中间前体	成熟 rRNA
HeLa 细胞	45S	→41S—┬—32S→ └—20S→	28 S 18 S
L 细胞	45S	→41S—┬—3.6S→32S→ └—→	28 S 18 S
爪蟾	40S	→38S—┬—34S—30S→ └—→	28 S 18 S
鼠肝	45S	→37.5S→36S—┬—32S—29.5S→ └—21.5S→	28 S 18 S
Novikoff 腹水瘤	45S	—┬—35S→n28S* └—23S→	28 S 18 S
昆虫**	38S	—┬—30S→ └—23S→	28 S 18 S
植物**	41S	→38S—┬—26S→ └—20S→	25 S 18 S
酵母	35S	—┬—29S→ └—20S→18S→	25 S 17 S
粘菌	36S	—┬—30S→28S→ └—19S→	26 S 17 S

* 核小体中的 28SRNA。
** 为某些昆虫和植物来源，不同来源加工过程是不同的。

对真核生物 rRNA 前体的加工酶的了解不如原核生物，它大概由类似 RNaseIII 以及其他内切核酸酶所完成。少数 rRNA 前体（主要在 28S RNA 中）存在插入顺序（见表 5-8）对于四膜虫 rRNA 前体加工的研究指出，接剪过程是在不存在酶蛋白情况下自我完成的(见第 172 页)；也有实验指出，某些小分子细胞核 RNA(snRNA) 如 U3、7S-1、8S 等在 rRNA 前体加工中起作用(见第八章)。

真核生物 rRNA 中也含有较多的甲基化成分，其中核糖的 2′-甲基化占甲基化成分的 80% 以上，如 HeLa 细胞 28S rRNA 有 65 个 2′-o-甲基核苷和 5 个甲基化碱基，18S rRNA 则含 40 个糖的甲基和 6 个碱基上甲基。用甲基标记的甲硫氨酸培养 HeLa 或 L

细胞，早期差不多全部同位素集中在核仁的 45S RNA 中，其后扩散至 32S 以及 28S RNA，与用标记尿苷培养得到完全相同的结果。利用 2′-甲基化核苷酸键的碱稳定性，比较带放射性的碱稳定二核苷酸种类观察到，32S RNA 中的与 28S RNA 相当，45S 的则等于 28S 与 18S 两者的总和。这些结果不仅支持上述加工过程的确定，同时指出，差不多所有 2′-羟基的甲基化是在 45S RNA 前体阶段进行的，表明修饰在剪切之前。同样，rRNA 中的 Ψ 也是在 45S 前体上由 U 异构化生成。另外在酵母 17S rRNA 和 HeLa 细胞 18S rRNA 中发现有一高度修饰成分 $m^1ncp^3\Psi$（见第一章），它的生物合成经过三个步骤：(i) 假尿苷化，(ii) 甲基化和 (iii) 导入 3-氨基-3-羧丙基。结果指出，第 (i)、(ii) 两步在早期前体中进行，第 (iii) 步则在核糖体接近成熟时的 18S 中间前体中反应。17—18S rRNA 的 3′ 端部位具有通用同源顺序 $-m^6_2A-m^6_2A-$，以及含其他修饰碱基如 m^6A、m^7G 等，对于它们的修饰过程了解很少。



前面提到，真核生物 5S rRNA 基因的转录与其他 rRNA 基因（构成 45S 的大前体）不是联在一起的。在爪蟾的卵早期发生过程中，5S RNA 大量堆积，比 18S RNA 和 28S RNA 大 100 倍。低浓度放线菌素 D 抑制 L 细胞中核仁的转录，但仍允许 5S RNA 继续生成。与 45S RNA 前体不同，5S RNA 却与 tRNA 相似，由 RNA 聚合酶 III 转录，大概 5S RNA 基因与 tRNA 基因具有相同的转录终止信号。由于同位素短时期标记能很快直接进入 5S RNA，因而可能 5S RNA 的原始转录产物不需进一步加工。真核生物 5S RNA 的 5′ 端通常为 pppG，也有 pppA 末端，至少说明转录产物没有前导顺序，不需进行剪切。

核糖体的生物合成

以上我们只是“孤立地”讨论 rRNA 的生物合成问题，其实，rRNA 的转录与加工是以核蛋白方式进行的，即 rRNA 前体与核糖体蛋白质结合成核蛋白颗粒，然后进行加工。细菌新合成的核糖体蛋白质一般比成熟的核糖体蛋白质稍大一点，切去其 N 端甲酰甲硫氨酸及其他几个氨基酸后，则成为成熟的核糖体蛋白质。真核细胞质的核糖体蛋白质则

是在细胞质中合成后进入细胞核,置换核糖体颗粒上的核仁蛋白质,形成新生的核糖体。

核糖体的体内装配首先由 rRNA 前体与核糖体蛋白质结合,随后可能通过不断添加核糖体蛋白质并进行 rRNA 前体加工,产生未成熟的核糖体亚基,最后经过必要的修饰(包括 RNA 和蛋白质修饰)才成为活性亚基,这个成熟过程无论对原核生物或真核生物都是必需的。

大肠杆菌核糖体生成过程见示意图 5-32。其中符号 P 代表 RNA 前体,m 表示成熟 rRNA,图中也以沉降常数 S 示出成熟过程核蛋白颗粒的大小,但没有指出 rRNA 的加工过程。

图 5-32 大肠杆菌核糖体的生物合成

真核生物细胞质核糖体的 RNA 在核仁中合成,新生的 rRNA 前体与核仁蛋白质形成蛋白颗粒,核仁蛋白质然后被来自细胞质的核糖体蛋白质所置换。颗粒中的 rRNA 前体经过逐步加工,并进入核质和通过核膜达到细胞质,这时可能也象原核生物那样有一个成熟过程,包括 rRNA 的最后加工和某些修饰反应(但大部分修饰是在更大前体水平上进行的,见第 175 页)。HeLa 细胞核糖体生物合成可以图 5-33 表示。圆圈代表核蛋白颗粒,圈外代表颗粒的沉降常数,圈内为 rRNA 的沉降常数。

图 5-33 HeLa 细胞核糖体的生物合成

5S(?) 表示 5S RNA 的生成和成熟过程不清楚

(图中 58S 应为 5.8S)

真核细胞器如线粒体、叶绿体具有自己独特的核糖体,同时组成一个区别于细胞质的蛋白质生物合成体系(见第八章)。

六 tRNA 的转录与加工

tRNA 的结构与功能已在第三章和第六章进行了讨论。和其他种类 RNA 一样,tRNA

的生物合成是以 tDNA 为模板，被 RNA 聚合酶(真核生物为 RNA 聚合酶 III)转录产生 tRNA 前体，然后经过加工得到成熟的 tRNA 分子。上节提到，一些 tRNA 基因是插在原核生物 rRNA 基因之中的，因而它们与 rRNA 一起转录，并大致同时成熟。

tRNA 基因（tDNA）与 tRNA 前体

tRNA 的种类远较 rRNA 为多，包括同功受体在内一般原核生物有三、四十种，而真核生物则有五、六十种或更多。在原核生物中，一些 tRNA 基因是与 rRNA 基因联接在一起的，包括间隔区中的 tRNA 基因和末端区的 tRNA 基因(表 5-11)。如大肠杆菌 K12 有七个 rRNA 基因簇，它的 16S rRNA 和 23S rRNA 基因之间都插有 tRNA 基因，其中三个编码 $tRNA^{Ile}$ 和 $tRNA^{Ala}_{1B}$，另外四个编码 $tRNA^{Glu}_{2}$；另外有几个 tRNA，如 $tRNA^{Asp}$、$tRNA^{Trp}$ 和 $tRNA^{Thr}$ 则由 rRNA 基因簇中远离启动子的末端顺序转录(表 5-11)。

表 5-11 大肠杆菌 K12 中和 rRNA 操纵子相联的 tRNA 基因

rRNA 操纵子	染色体位置	间隔区 tRNA	末端区 tRNA
rrnA	85	$tRNA^{Ile}_{1}$,$tRNA^{Ala}_{1B}$	
rrnB	88	$tRNA^{Glu}_{2}$	
rrnC	83	$tRNA^{Glu}_{2}$	$tRNA^{Asp}_{1}$,$tRNA^{Trp}$
rrnD	71	$tRNA^{Ile}_{1}$,$tRNA^{Ala}_{1B}$	$tRNA^{Thr}$
rrnE	89	$tRNA^{Glu}_{2}$	
组 I	–	$tRNA^{Ile}_{1}$,$tRNA^{Ala}_{1B}$	$tRNA^{Asp}_{1}$
组 II	–	$tRNA^{Glu}_{2}$	

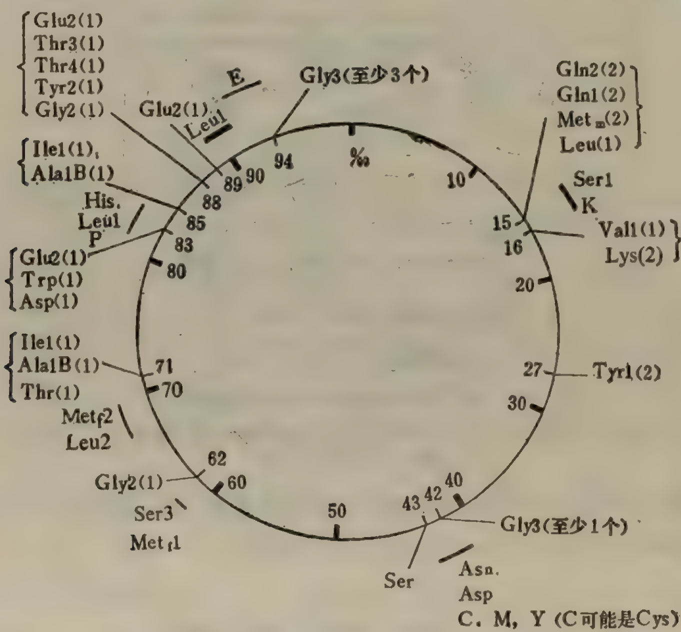


图 5-34 大肠杆菌基因图上 tRNA 基因的位置

图中也标明同功受体基因，括弧内是基因拷贝数，圆圈内数字以分表示，C、E、K、M、P 和 Y 是未确定的 tRNA 基因。

此外，tRNA 基因大都以基因簇存在，构成多顺反子转录单位。大肠杆菌基因图上 tRNA 基因的已知位置见图 5-34，已经测定了大肠杆菌中的约 60 个 tRNA 基因，若考虑摆动假说，它约为翻译 61 个密码子所需要的二倍(参阅遗传密码一章)。

1981 年鉴定了一个大肠杆菌操纵子，其中含七个 tRNA 基因：Met_m-Leu-Gln1-Gln1-Met_m-Gln2-Gln2（基因图的 15 分处），各 tRNA 基因间被二级结构所隔开，推想为 RNA 加工酶的识别点。

T4 噬菌体中稳定 RNA 转录单位长约 3,000 个核苷酸，其中含一个蛋白质基因和二个 RNA 基因簇：一个簇有 $tRNA^{Arg}$ 和二个未知功能的稳定 RNA，另一个簇从 5′ → 3′ 编码七个 tRNA：Gln-Leu-Gly-Pro-Ser-Thr-Ile(图 5-35)。

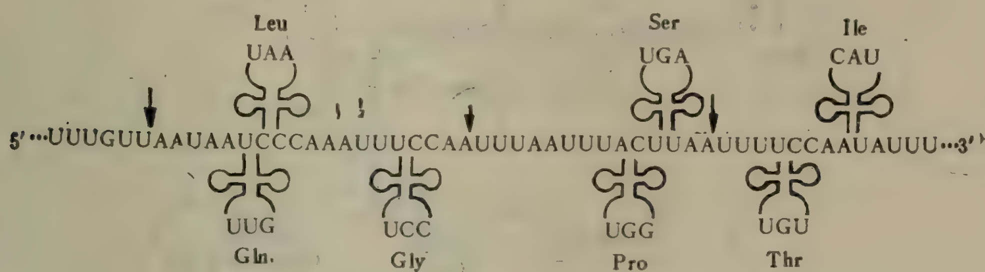


图 5-35 T4 tRNA 基因中一个含七个 tRNA 基因簇的转录产物

由于加工紧随着转录之后，转录没有终止或开始不久加工就已进行(原核生物 rRNA 合成也有类似情况，如上述)，所以在野生型细菌很难分离出大的转录产物。企图从 T4 感染细胞分离单一的多顺反子转录产物也没有成功。图 5-36 示出双顺反子 tRNA 基因转录与加工过程，一种为单顺反子产物，另一种是双顺反子产物，后者常在加工酶缺陷型菌中分离得到；两种产物都需要进一步加工才产生成熟 tRNA 分子。

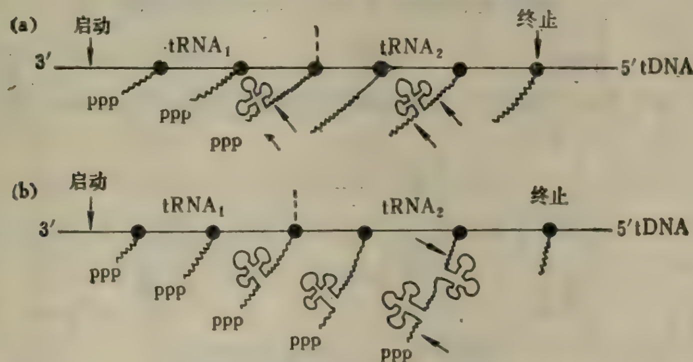


图 5-36 tDNA 的转录与加工关系示意图

● 表示 RNA 聚合酶；↑为 RNA 加工酶的剪切位置；～～代表前导顺序或间隔区。(a) 为单顺反子转录产物；(b) 为双顺反子转录产物。

目前，有三类 tRNA 前体已被分离纯化，并对其加工过程进行了研究。它们是(i)单顺反子转录产物，(ii) 双(多)顺反子转录产物以及 (iii) 在真核生物中发现的有插入顺序的转录产物。

最早分离并测定其一级结构的 tRNA 前体是大肠杆菌 SU_3^+ $tRNA^{Tyr}$ 前体，与成熟的 $tRNA^{Tyr}_{SU_3^+}$ 比较(结构见第六章)在 5′端从 pppG 开始多了 41 个核苷酸，3′端也多了几个核苷酸，同时前体中不存在修饰成分，但分子内已含有 3′-CCA 末端。

在缺乏加工酶 RNaseP 的变异株中，可分离出一些 tRNA 前体，已测定序列的单顺反子前体有 $tRNA^{Phe}$、$tRNA_2^{Glu}$ 和 $tRNA^{Asp}$ 前体等，这些前体也都在 5′ 端含附加顺序，但 3′ 端是成熟的。$SU_3^+tRNA^{Tyr}$ 前体的结构见图 5-37。

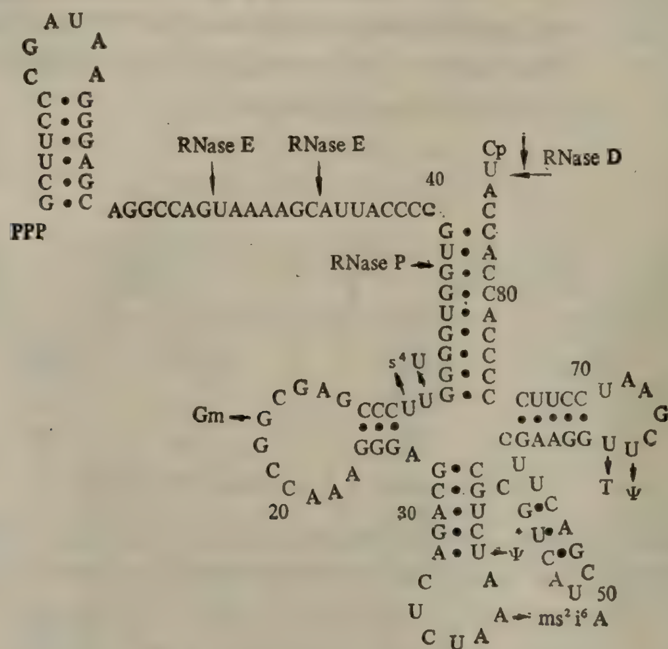


图 5-37　$tRNA^{Tyr}_{SU_3^+}$ 前体的三叶草结构

箭号处表示与成熟 $tRNA^{Tyr}_{SU_3^+}$ 不同或需剪切部位

从大肠杆菌中也分离鉴定了一个双顺反子前体 $tRNA_2^{Gly}$-$tRNA^{Thr}$，除了在 5′和 3′端含附加顺序外，两个 tRNA 间存在 6 核苷酸的间隔区(图 5-38)。从图中可以看到，前体

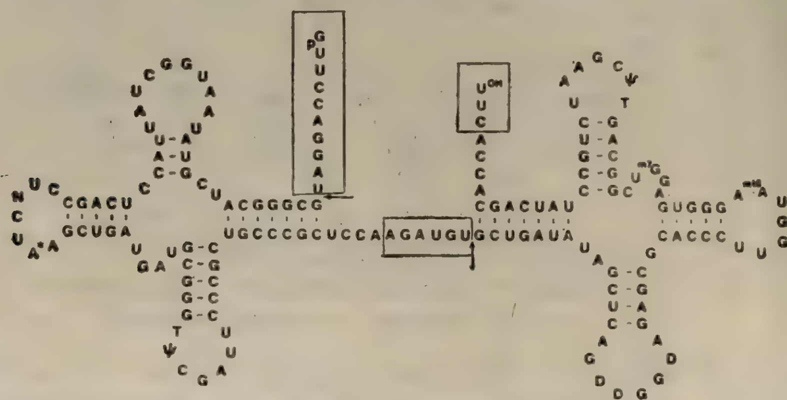


图 5-38　大肠杆菌 $tRNA_2^{Gly}$-$tRNA^{Thr}$ 双顺反子前体

方框表示需要被切除部分

分子中已经存在成熟 tRNA 的修饰成分。已知被研究的所有大肠杆菌 tRNA 前体都含有 tRNA 的 3′-CCA 顺序,但其他生物的 tRNA 前体链中,并不都含有 3′-CCA。

从 T4 噬菌体的转录产物中也已分离出单顺反子 tRNA 前体以及双顺反子前体,如 $tRNA^{Pro}$-$tRNA^{Ser}$, $tRNA^{Gln}$-$tRNA^{Leu}$,$tRNA^{Thr}$-$tRNA^{Ile}$。在这些双体的 tRNA 前体中,其 5′ 和 3′ 端两侧均有长短不一的附加顺序;此外两个 tRNA 之间也为数个或者一个核苷酸所隔开;在 T4 双体 tRNA 中,有的 tRNA 已经存在 3′-CCA 末端,另一些则不是。值得注意的是,在这些双体分子中含有大部分修饰成分,表明修饰在剪切之前出现(图 5-39)。

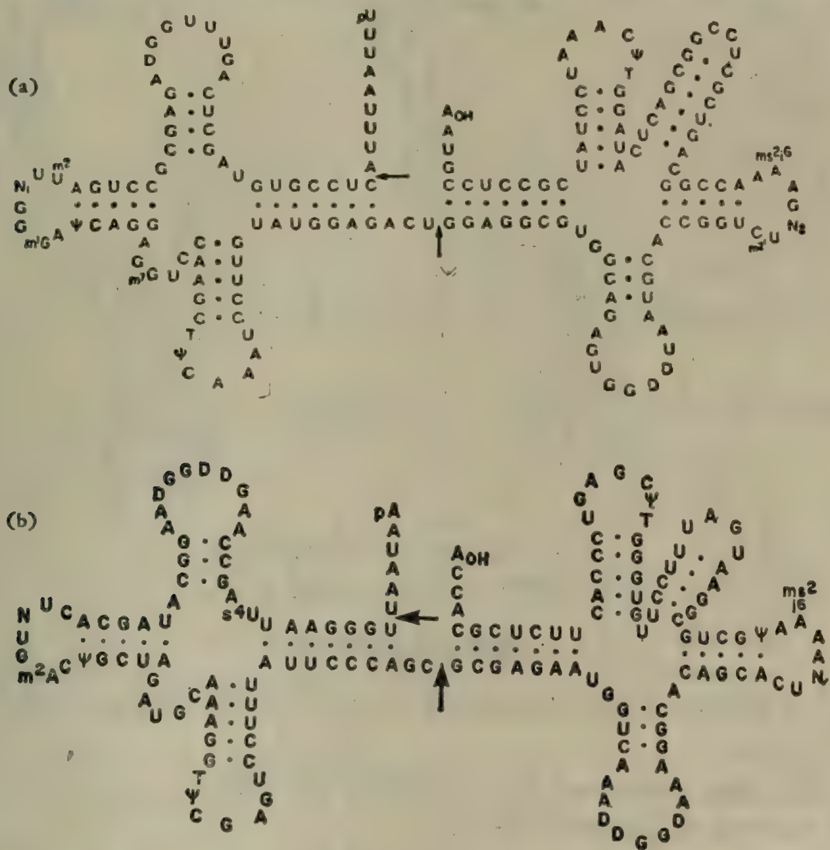


图 5-39 T4 $tRNA^{Pro}$-$tRNA^{Ser}$(a) 和 $tRNA^{Gln}$-$tRNA^{Leu}$(b) 前体分子

真核生物 tRNA 基因数目比原核生物大得多,如大肠杆菌有大约 60 个 tRNA 基因,酵母则有 320—400 个,果蝇 750 个,而爪蟾差不多有 8,000 个,其中每种 tRNA 基因约 200 个拷贝。真核生物 tRNA 基因大概也是以簇存在的,并且为间隔区所分开。但目前分离得到的 tRNA 前体都是单顺反子的。用聚丙烯酰胺凝胶电泳可以从脉冲标记的 HeLa 细胞、家蚕、酵母中分离出较多的介乎 4S 和 5S 之间(约 4.5S)的 tRNA 前体分子,这些分子与成熟 4S 比较,主要是 5′ 和 3′ 端存在附加顺序。5′ 端含核苷-5′-三磷酸,表明是原始转录产物,同时在前体中已经存在部分修饰成分。

1978 年以来从酵母变异株中分离出含插入顺序的 tRNA 前体,所有这些前体都在反密码臂处插入十几至三十几个核苷酸,图 5-40 列出的 tRNA 前体中,$tRNA^{Trp}$ 和 $tRNA_3^{Leu}$ 的插入顺序部分还出现一个新的茎环结构。在其他一些来源的真核 tRNA 前体

中也发现有插入顺序存在，虽然它不如 mRNA 前体那么普遍，但至少表明，插入顺序是真核生物与原核生物区别的一个重要标记。

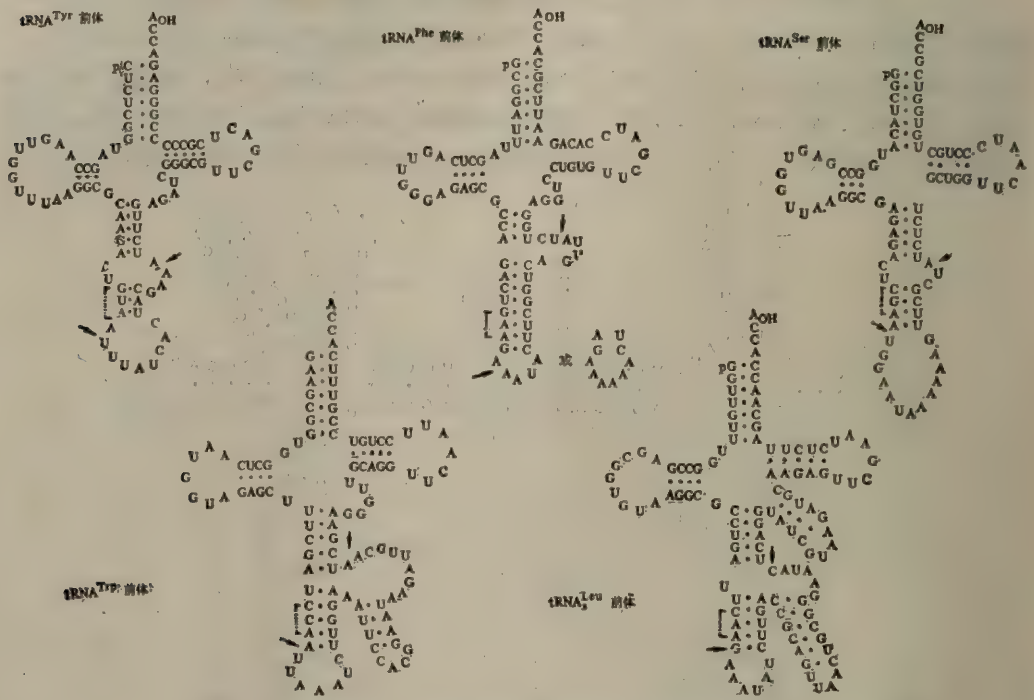


图 5-40　五个酵母 tRNA 前体的二级结构

其中↑为剪接点，⊏为成熟 tRNA 的反密码子。

tRNA 前体的加工

tRNA 前体的加工内容具有与其他种类 RNA 的共性，同时也有本身的特点，主要包括以下三个方面内容:

(1) 剪切、修剪 (trimming、clipping) 和剪接 (splicing);

(2) 如果需要，在 3′ 端添加 CCA_{OH};

(3) 核苷修饰。

所有这些反应都是在酶催化下进行的，总的来说，tRNA 加工所需用的酶比 mRNA 或 rRNA 加工都要多，特别是多顺反子 tRNA 基因转录产物的加工更为复杂，往往需要多个酶的协同作用。

剪切、修剪和剪接

tRNA 剪切酶是一类主要的加工酶，属内切核酸酶性质，它在 tRNA 前体两旁附加顺序处水解，以得到一定大小的 tRNA 分子。大肠杆菌 RNase P 是特异剪切差不多所有大肠杆菌及其噬菌体 tRNA 前体的 5′ 旁侧顺序的，生成成熟 tRNA 的 5′ 末端，因此这个 5′-内切核酸酶也被叫作 tRNA 5′ 成熟酶，比较了几个 tRNA 前体的 RNase P 作用位点指出，RNase P 的专一性来看与其剪切点周围的一级结构以至二级结构并无直接关系。在体外，人工除去 $tRNA^{Tyr}_{SU_3^+}$ 前体中前导顺序 41 个核苷酸中的 32 个核苷酸，并不影

响 RNase P 的作用；由于人工合成 $tRNA^{Pro}$-$tRNA^{Ser}$ 前体中 RNase P 剪切点周围 4 和 14 核苷酸长的寡核苷酸片段不能作为 RNase P 的底物，因此认为 RNase P 的严格的专一性可能与 tRNA 前体分子的特殊构型有关。RNase P 也能加工某些非 tRNA 前体分子，如大肠杆菌 4.5S RNA 前体，φ80 噬菌体编码的另一 RNA。经过分析，发现这些 RNA 也具有类似 tRNA 的三叶草结构。3′ 端无 CCA_{OH} 的 tRNA 前体不能作为 RNase P 的底物，或不是好的底物。用化学方法或酶法除去大肠杆菌 $tRNA_1^{Tyr}$ 的 3′-CCA_{OH} 一个或二个末端核苷酸，即明显影响它们被 RNase P 的水解速度。

1978 年 Altman 等发现大肠杆菌 RNase P 是一种不寻常的酶分子，它由至多 23% 蛋白质和至少 77% RNA 所组成，沉降密度为 1.71g/cm³(CsCl 中) 和 1.55g/cm³(Cs_2SO_4 中)，是一个特征的 RNA-蛋白质复合物，其中多肽分子量约 17,500d，RNA 分子量为 120,000d，约为 360 核苷酸长。用蛋白水解酶、固定化 RNase A 或依赖 Ca^{2+} 的小球菌核酸酶处理则 RNase P 失活；拆开蛋白质和 RNA，两者不具活性，重组则活性恢复。1980 年分离得到的枯草杆菌 RNase P 也是一个由 RNA 和蛋白质组成的酶。RNase P 中存在 RNA 组分表明，它可能是一个识别亚基，具有仅与 tRNA 及结构类似的 RNA 相互作用的特定顺序与构象。从其他许多来源如酵母、昆虫、哺乳动物中也都发现有类似 RNase P 活性。

除了 RNase P 作为 5′-内切核酸酶外，加工 tRNA 前体至少还需要一个 3′-内切核酸酶(如 RNase F)，它从较大前体切下 3′端带有不太长的核苷酸顺序。用于 rRNA 加工的 RNase III 和 RNase E 在 tRNA 加工中可能也起作用，但 RNase III 对于体内 tRNA 成熟并非绝对需要，而 RNase E 剪切仅用于很少数 tRNA 前体中，这两个酶的功能大概是使大的新生的 tRNA 转录产物转变为小的产物，后者是 RNase P 更有效的底物。

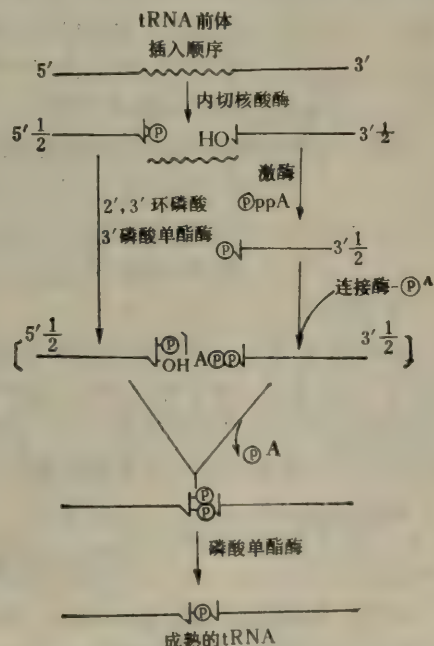
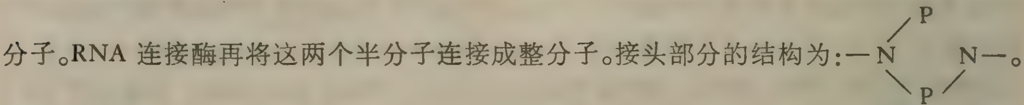


图 5-41 tRNA 前体中插入顺序的切除
(1/2代表半分子)

为了使 tRNA 分子得到成熟的 3′端，还需要有一个外切酶进行修剪，从 3′ 端逐个切去附加的顺序，以暴露出 tRNA3′ 端。这个修剪酶是 RNaseD，分子量为 38,000d，为单一亚基的蛋白质。它不同于另一 3′ 外切酶 RNase II，RNase D 具更为严格的选择活性，因此 RNase D 是一个体内的 3′-tRNA 成熟酶。

tRNA˙ 前体中插入顺序的除去，是靠一非常复杂的酶系来完成的。现已从酵母中分离得到用于剪接反应的内切核酸酶和 RNA 连接酶。内切核酸酶可将 tRNA 前体加工成带有 2′、3′ 环磷酸的 5′ 半分子和 5′ 羟基的 3′ 半分子，同时释放出完整的插入顺序。它是一种单一的内切核酸酶，可识别所有 tRNA 前体的共同结构特征，加工各种不同的同源或异源的 tRNA 前体。5′ 半分子在 2′、3′ 环磷

酸 3′ 单酯酶作用下成带 2′磷酸的 5′半分子。3′半分子在激酶作用下生成带5′ 磷酸的 3′ 半分子。RNA 连接酶再将这两个半分子连接成整分子。接头部分的结构为：

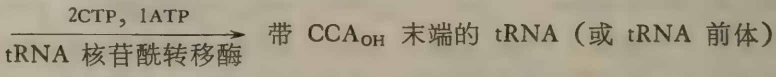


由于产物中 2′ 磷酸不能被一般的磷酸单酯酶除去，故推测还存在一种新的特异性的磷酸单酯酶。 反应机制如图 5-41 所示。以前称之谓剪接酶的，就是由上述各酶所组成的酶系。不过有证据表明，剪接反应的机制是多种多样的，可能有几种反应途径。

CCA 添加

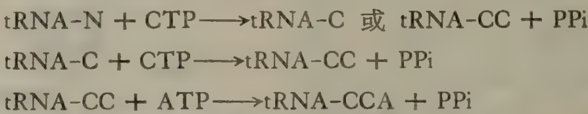
缺乏或具有不完全的 3′-CCA_{OH} 末端的 tRNA 是没有活性的。 有的 tRNA 基因中带有—CCA 模板顺序，而另一些则无，因此经过剪切或修剪后的 tRNA，其 3′ 端可能存在但也可能没有 3′-CCA_{OH}，对于后者，需要在 tRNA 核苷酰转移酶催化下进行末端加成：

缺 CCA_{OH} 末端的 tRNA（或 tRNA 前体）



$\xrightarrow[\text{tRNA 核苷酰转移酶}]{\text{2CTP, 1ATP}}$ 带 CCA_{OH} 末端的 tRNA（或 tRNA 前体）

tRNA 核苷酰转移酶广泛分布于所有细胞、细胞器和病毒中，它对供体的要求是非常专一的，除 CTP 和 ATP 外，其他 NTP 不能作为供体。体外实验表明，若只存在一种供体（CTP 或 ATP），也可催化下列反应：

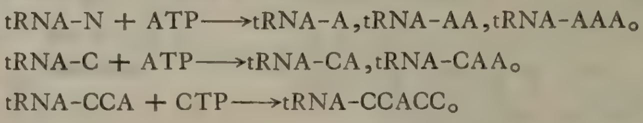


tRNA-N + CTP⟶tRNA-C 或 tRNA-CC + PPi

tRNA-C + CTP⟶tRNA-CC + PPi

tRNA-CC + ATP⟶tRNA-CCA + PPi

同时，在某些特殊情况下，该酶也能催化不正常的合成，如



tRNA-N + ATP⟶tRNA-A，tRNA-AA，tRNA-AAA。

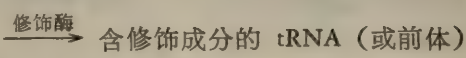
tRNA-C + ATP⟶tRNA-CA，tRNA-CAA。

tRNA-CCA + CTP⟶tRNA-CCACC。

核苷修饰

成熟的 tRNA 分子中含有许多修饰成分，不同 tRNA 所含修饰成分种类和数目都各不相同（详见第一章和第三章）。虽然修饰成分在 RNA 中的生物功能仍了解很少，但对其生物合成已有不同程度的了解。核苷修饰问题是 RNA 特别是 tRNA 加工的主要内容。tRNA 的修饰是在多核苷酸链上进行的，由一类统称为修饰酶所催化：

缺修饰成分的 tRNA（或前体）



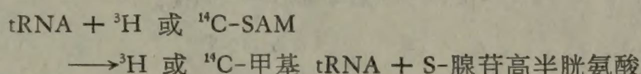
$\xrightarrow{\text{修饰酶}}$ 含修饰成分的 tRNA（或前体）

tRNA 修饰酶具有高度专一性，DNA、rRNA（可能有例外）、多核苷酸或单核苷酸、核苷都不能作为修饰酶的底物。由于 tRNA 中修饰成分繁多，因而修饰酶种类很多，一般来说，每种修饰核苷都有催化合成本身的修饰酶。

tRNA 甲基转移酶（简称 tRNA 甲基化酶）是研究较早也较详细的一类修饰酶，它不同于前面的 mRNA 甲基转移酶，与甲基化 rRNA 分子的酶也可能不相同。近年来，由于酶分离纯化技术的进展，已经得到对碱基以及排列顺序高度专一的 tRNA 甲基化酶，如 tRNA（腺嘌呤-1）甲基化酶催化 tRNA 中特定位置的 A → m^1A，tRNA（鸟嘌呤-7）甲

基化酶可使 tRNA 中第55位的 $G \rightarrow m^7G$，其他如催化 tRNA 中生成 m^1G、m^2G、m^5C 以及 rT 的酶也已高度纯化。有的甲基化酶则能同时催化二种以上核苷，可能是不同甲基化酶的混合物。tRNA 甲基化酶也具有严格的顺序要求，如从酵母分离得二个 tRNA（鸟嘌呤-1）甲基化酶，它们能分别甲基化 tRNA 中的 G_{19} 和 G_{43} 生成 m^1G_{19} 和 m^1G_{43}。从枯草杆菌分离的 tRNA（腺嘌呤-1）甲基化酶能催化 tRNA 中的···$YAAGGCm_2^2G \longrightarrow YAm^1AGGCm_2^2G$；而鼠肝中得到的同一酶则使 $GT\phi CGAAU \rightarrow GT\phi CGAm^1AU$。

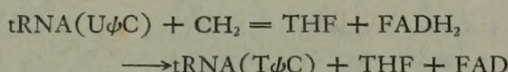
差不多所有 tRNA 甲基化酶的甲基供体均为 S-腺苷甲硫氨酸（SAM），使用放射性标记的 SAM 进行实验，可以简便地测定甲基化酶活性：



tRNA + ^{3}H 或 ^{14}C-SAM

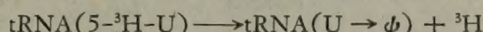
$\longrightarrow$ ^{3}H 或 ^{14}C-甲基 tRNA + S-腺苷高半胱氨酸

但也有报道，在某些革兰氏阳性菌中分离出 tRNA（尿嘧啶-5）甲基化酶（即 rT 甲基化酶），所用供体不是 SAM，而是5,10-甲叉四氢叶酸，同时以 $FADH_2$ 作还原剂，反应如下：



$$tRNA(U\phi C) + CH_2 = THF + FADH_2$$
$$\longrightarrow tRNA(T\phi C) + THF + FAD$$

被广泛进行研究的 tRNA 修饰酶还有：tRNA 异戊烯转移酶（tRNAΔ^2 异戊烯合成酶），i^6A 上的异戊烯侧链的供体是 $\triangle^2$-异戊烯焦磷酸；tRNA t^6A 合成酶，在碳酸氢盐存在下，Thr 进入 tRNA，生成 t^6A；催化 yw(Y)生成的酶，Y的第三环来自 Met 中的3-氨基-3-羧丙基；tRNA：鸟嘌呤转糖苷酶，催化 tRNA 中Q的生成，即Q碱基的进入是通过 N-糖苷键断裂进行的；tRNAϕ 合成酶，至少有二个酶分别催化反密码区上ϕ的生成和其他区（包括 TϕC 区）上ϕ的生成，实质上酶催化异构化作用：



$$tRNA(5\text{-}^3H\text{-}U) \longrightarrow tRNA(U \rightarrow \phi) + {}^3H$$

tRNA 硫转移酶不止一种，催化 S^4U 以及其他含硫的嘧啶化合物的生成，供体是 Cys 或 β-巯基丙酮酸；催化 tRNA 中生成 I、D、ncp^3U 的酶活性也已发现。

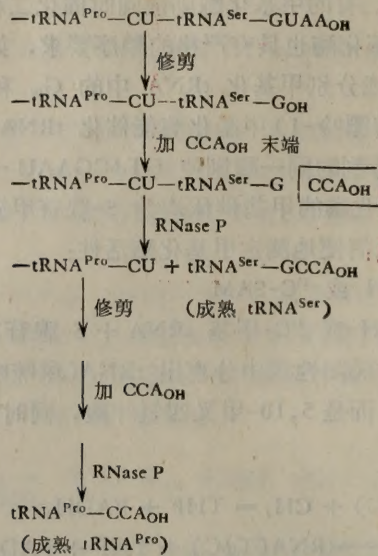
关于 tRNA 中修饰核苷的生物合成详见本“丛书”代谢部分的核酸代谢一章。

下面具体介绍几个 tRNA 前体的加工过程。图5-37是 $tRNA^{Tyr}_{SU_3^+}$ 前体的三叶草结构，图中也示出剪切和修剪部位和所需的加工酶：RNase P、RNase E 和 RNase D；修饰成分的生成包括 U_8 与 $U_9 \rightarrow S^4U$、$G_{17} \rightarrow Gm$、A_{38} 通过 i^6A 生成 ms^2i^6A、U_{40} 转变为 ϕ、U_{63} 和 U_{64} 分别变成 T 和 ϕ 等。

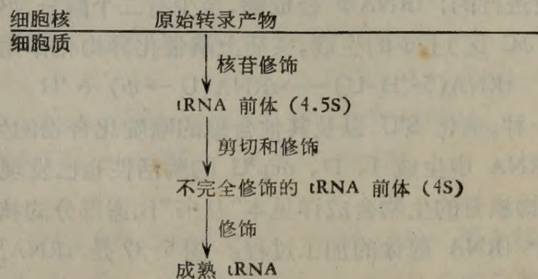
较为复杂的例子是前面提到T4噬菌体中存在由七个 tRNA 组成的基因簇（图5-35），但一般分离不到完整的转录产物，因为感染后很快就出现剪切产物，并且得到成熟 tRNA。但在早期也分离出三个 tRNA 双顺反子前体（见第181页）：$tRNA^{Pro}$-$tRNA^{Ser}$、$tRNA^{Thr}$-$tRNA^{Ile}$ 以及 $tRNA^{Gln}$-$tRNA^{Leu}$，后者含量很少，因为它被迅速剪切为成熟的 $tRNA^{Leu}$ 和 $tRNA^{Gln}$ 前体。一种解释是，在 $tRNA^{Gln}$-$tRNA^{Leu}$ 双体中的 $tRNA^{Leu}$ 的3′端存在 CCA_{OH} 末端（见图5-39b）而使 RNase P 表现出较高活性（见第183页），而在 $tRNA^{Pro}$-$tRNA^{Ser}$ 和 $tRNA^{Thr}$-$tRNA^{Ile}$ 前体中，由于3′端无 CCA_{OH}（图5-39a），RNase P 不能使其3′部分的 tRNA 的5′端成熟，故双体堆积。而只有当这些双体分子的3′端在 tRNA 核苷酰转移酶催化下加上 CCA_{OH}，RNase P 才作用于双体，产生 $tRNA^{Ser}$、$tRNA^{Ile}$ 以及



$tRNA^{Pro}$ 前体和 $tRNA^{Thr}$ 前体。以 $tRNA^{Pro}$-$tRNA^{Ser}$ 前体加工为例:



从以上举的例子可以看到,在加工过程中核苷修饰可能在剪切和修剪之后,但也有的在剪切和修剪之前,没有规律性。真核生物 tRNA 前体加工也具有类似情况。真核生物 tRNA 前体约 100 核苷酸长,只含一个单一 tRNA 顺序,同时存在较少的修饰成分,显然它不是 tDNA 的原始转录产物。这表明核苷修饰可能有先有后,分两阶段进行。



含插入顺序的 tRNA 前体的剪接部位见图 5-40,详细过程见图 5-41。

参　考　文　献

[1] Transfer RNA: Biological Aspect (eds. Söll, D., Abelson, J. N. and Schimmel, P. R.) 1980, Cold Spring Harbor Laboratory.

[2] The Enzymes (ed Boyer, P. D) 1982, Vol. XV, Nucleic Acids, Part B, 3rd ed.

[3] Gegenheimer, P. and Apirion, D. 1981, Processing of Procaryotic RNA, in *Microbiological Reviews*, **45**, 502.

[4] Big Problems Faced in RNA Processing (Research News), 1982, *Science*, **216**, 1208.

[5] Yanofsky, C., 1981, *Nature*, 289, 751.

[6] Platt, T., 1981, *Cell*, **24**, 10.

[7] RNA Can Be A Catalyst (Research News), 1982, *Science*, **218**, 872.

[8] Busch, H. et al., 1982, SnRNPs, And RNA Processing, *in Ann. Rev Biochem.* (eds. Snell, E. E. et al.) **51**, 617, Annual Rev. Inc. USA.

统一书号：13031·333
定 价： 3.00 元
本社书号：4055·13-10
科技新书目：129-33